

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

AD

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : C12N 15/12, C07K 14/72, G01N 33/50, 33/566	A1	(11) International Publication Number: WO 00/22129 (43) International Publication Date: 20 April 2000 (20.04.00)
(21) International Application Number: PCT/US99/23938 (22) International Filing Date: 12 October 1999 (12.10.99) (30) Priority Data: 09/170,496 13 October 1998 (13.10.98) US (63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application US 09/170,496 (CIP) Filed on 13 October 1998 (13.10.98) (71) Applicant (for all designated States except US): ARENA PHARMACEUTICALS, INC. [US/US]; 6166 Nancy Ridge Drive, San Diego, CA 92121 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): BEHAN, Dominic, P. [GB/US]; 11472 Roxboro Court, San Diego, CA 92131 (US). CHALMERS, Derek, T. [GB/US]; 347 Longden Lane, Solana Beach, CA 92075 (US). LIAW, Chen, W. [US/US]; 7668 Salix Place, San Diego, CA 92129 (US).		(74) Agents: MILLER, Suzanne, E. et al.; Woodcock Washburn Kurtz Mackiewicz & Norris LLP, 46th floor, One Liberty Place, Philadelphia, PA 19103 (US). (81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
(54) Title: NON-ENDOGENOUS, CONSTITUTIVELY ACTIVATED HUMAN G PROTEIN-COUPLED RECEPTORS (57) Abstract <p>Disclosed herein are constitutively activated, non-endogenous versions of endogenous human G protein-coupled receptors comprising (a) the following amino acid sequence region (C-terminus to N-terminus orientation) and/or (b) the following nucleic acid sequence region (3' to 5' orientation) transversing the transmembrane-6 (TM6) and intracellular loop-3 (IC3) regions of the GPCR: (a) P¹ AA₁₅ X and/or (b) P_{codon} (AA-codon)₁₅ X_{codon}, respectively. In a most preferred embodiment, P¹ and P_{codon} are endogenous proline and an endogenous nucleic acid encoding region encoding proline, respectively, located within TM6 of the non-endogenous GPCR; AA₁₅ and (AA-codon)₁₅ are 15 endogenous amino acid residues and 15 codons encoding endogenous amino acid residues, respectively; and X and X_{codon} are non-endogenous lysine and a non-endogenous nucleic acid encoding region encoding lysine, respectively, located within IC3 of the non-endogenous GPCR. Because it is most preferred that the non-endogenous human GPCRs which incorporate these mutations are incorporated into mammalian cells and utilized for the screening of the candidate compounds, the non-endogenous human GPCR incorporating the mutation need not be purified and isolated <i>per se</i> (i.e., these are incorporated within the cellular membrane of a mammalian cell), although such purified and isolated non-endogenous human GPCRs are well within the purview of this disclosure.</p>		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

**NON-ENDOGENOUS, CONSTITUTIVELY ACTIVATED
HUMAN G PROTEIN-COUPLED RECEPTORS**

The benefits of commonly owned U.S. Serial Number 09/170,496, filed October 13, 1998, U.S. Serial Number 08/839, 449 filed April 14, 1997 (now abandoned),
5 U.S. Serial Number 09/060,188, filed April 14, 1998; U.S. Provisional Number 60/090,783, filed June 26, 1998; and U.S. Provisional Number 60/095,677, filed on August 7, 1998, are hereby claimed. Each of the foregoing applications are incorporated by reference herein in their entirety.

FIELD OF THE INVENTION

10 The invention disclosed in this patent document relates to transmembrane receptors, and more particularly to human G protein-coupled receptors (GPCRs) which have been altered such that altered GPCRs are constitutively activated. Most preferably, the altered human GPCRs are used for the screening of therapeutic compounds.

BACKGROUND OF THE INVENTION

Although a number of receptor classes exist in humans, by far the most abundant and therapeutically relevant is represented by the G protein-coupled receptor (GPCR or GPCRs) class. It is estimated that there are some 100,000 genes within the human genome, and of these, approximately 2% or 2,000 genes, are estimated to code for GPCRs. Of these, there are approximately 100 GPCRs for which the endogenous ligand that binds to the GPCR has been identified. Because of the significant time-lag that exists between the discovery of an endogenous GPCR and its endogenous ligand, it can be presumed that the remaining 1,900 GPCRs will be identified and characterized long before the endogenous ligands for these receptors are identified. Indeed, the rapidity by which the Human Genome Project is sequencing the 100,000 human genes indicates that the remaining human GPCRs will be fully sequenced within the next few years. Nevertheless, and despite the efforts to sequence the human genome, it is still very unclear as to how scientists will be able to rapidly, effectively and efficiently exploit this information to improve and enhance the human condition. The present invention is geared towards this important objective.

Receptors, including GPCRs, for which the endogenous ligand has been identified are referred to as "known" receptors, while receptors for which the endogenous ligand has not been identified are referred to as "orphan" receptors. This distinction is not merely semantic, particularly in the case of GPCRs. GPCRs represent an important area for the development of pharmaceutical products: from approximately 20 of the 100 known GPCRs, 60% of all prescription pharmaceuticals have been developed. Thus, the orphan GPCRs are to the pharmaceutical industry what gold was to California in the late 19th century – an opportunity to drive growth, expansion, enhancement and development. A serious drawback exists, however,

- 3 -

with orphan receptors relative to the discovery of novel therapeutics. This is because the traditional approach to the discovery and development of pharmaceuticals has required access to both the receptor *and* its endogenous ligand. Thus, heretofore, orphan GPCRs have presented the art with a tantalizing and undeveloped resource for the discovery of pharmaceuticals.

- 5 Under the traditional approach to the discovery of potential therapeutics, it is generally the case that the receptor is first identified. Before drug discovery efforts can be initiated, elaborate, time consuming and expensive procedures are typically put into place in order to identify, isolate and generate the receptor's endogenous ligand – this process can require from between 3 and ten years per receptor, at a cost of about \$5million (U.S.) per receptor. These time and financial
- 10 resources must be expended before the traditional approach to drug discovery can commence. This is because traditional drug discovery techniques rely upon so-called "competitive binding assays" whereby putative therapeutic agents are "screened" against the receptor in an effort to discover compounds that either block the endogenous ligand from binding to the receptor ("antagonists"), or enhance or mimic the effects of the ligand binding to the receptor ("agonists").
- 15 The overall objective is to identify compounds that prevent cellular activation when the ligand binds to the receptor (the antagonists), or that enhance or increase cellular activity that would otherwise occur if the ligand was properly binding with the receptor (the agonists). Because the endogenous ligands for orphan GPCRs are by definition not identified, the ability to discover novel and unique therapeutics to these receptors using traditional drug discovery techniques is not
- 20 possible. The present invention, as will be set forth in greater detail below, overcomes these and other severe limitations created by such traditional drug discovery techniques.

GPCRs share a common structural motif. All these receptors have seven sequences of between 22 to 24 hydrophobic amino acids that form seven alpha helices, each of which spans the

membrane (each span is identified by number, *i.e.*, transmembrane-1 (TM-1), transmembrane-2 (TM-2), etc.). The transmembrane helices are joined by strands of amino acids between transmembrane-2 and transmembrane-3, transmembrane-4 and transmembrane-5, and transmembrane-6 and transmembrane-7 on the exterior, or "extracellular" side, of the cell membrane (these are referred to as "extracellular" regions 1, 2 and 3 (EC-1, EC-2 and EC-3), respectively). The transmembrane helices are also joined by strands of amino acids between transmembrane-1 and transmembrane-2, transmembrane-3 and transmembrane-4, and transmembrane-5 and transmembrane-6 on the interior, or "intracellular" side, of the cell membrane (these are referred to as "intracellular" regions 1, 2 and 3 (IC-1, IC-2 and IC-3), respectively). The "carboxy" ("C") terminus of the receptor lies in the intracellular space within the cell, and the "amino" ("N") terminus of the receptor lies in the extracellular space outside of the cell. The general structure of G protein-coupled receptors is depicted in Figure 1.

Generally, when an endogenous ligand binds with the receptor (often referred to as "activation" of the receptor), there is a change in the conformation of the intracellular region that allows for coupling between the intracellular region and an intracellular "G-protein." Although other G proteins exist, currently, Gq, Gs, Gi, and Go are G proteins that have been identified. Endogenous ligand-activated GPCR coupling with the G-protein begins a signaling cascade process (referred to as "signal transduction"). Under normal conditions, signal transduction ultimately results in cellular activation or cellular inhibition. It is thought that the IC-3 loop as well as the carboxy terminus of the receptor interact with the G protein. A principal focus of this invention is directed to the transmembrane-6 (TM6) region and the intracellular-3 (IC3) region of the GPCR.

Under physiological conditions, GPCRs exist in the cell membrane in equilibrium between

- 5 -

two different conformations: an "inactive" state and an "active" state. As shown schematically in Figure 2, a receptor in an inactive state is unable to link to the intracellular signaling transduction pathway to produce a biological response. Changing the receptor conformation to the active state allows linkage to the transduction pathway (via the G-protein) and produces a biological response.

5 A receptor may be stabilized in an active state by an endogenous ligand or a compound such as a drug. Recent discoveries, including but not exclusively limited to modifications to the amino acid sequence of the receptor, provide means other than endogenous ligands or drugs to promote and stabilize the receptor in the active state conformation. These means effectively stabilize the receptor in an active state by simulating the effect of an endogenous ligand binding
10 to the receptor. Stabilization by such ligand-independent means is termed "constitutive receptor activation."

As noted above, the use of an orphan receptor for screening purposes has not been possible. This is because the traditional "dogma" regarding screening of compounds mandates that the ligand for the receptor be known. By definition, then, this approach has no applicability with
15 respect to orphan receptors. Thus, by adhering to this dogmatic approach to the discovery of therapeutics, the art, in essence, has taught and has been taught to forsake the use of orphan receptors unless and until the endogenous ligand for the receptor is discovered. Given that there are an estimated 2,000 G protein coupled receptors, the majority of which are orphan receptors, such dogma castigates a creative, unique and distinct approach to the discovery of therapeutics.

20 Information regarding the nucleic acid and/or amino acid sequences of a variety of GPCRs is summarized below in Table A. Because an important focus of the invention disclosed herein is directed towards orphan GPCRs, many of the below-cited references are related to orphan GPCRs. However, this list is not intended to imply, nor is this list to be construed, legally or

otherwise, that the invention disclosed herein is only applicable to orphan GPCRs or the specific GPCRs listed below. Additionally, certain receptors that have been isolated are not the subject of publications per se; for example, reference is made to a G Protein-Coupled Receptor database on the "world-wide web" (neither the named inventors nor the assignee have any affiliation with this site) that lists GPCRs. Other GPCRs are the subject of patent applications owned by the present assignee and these are not listed below (including GPR3, GPR6 and GPR12; *see* U.S. Provisional Number 60/094879):

Table A

Receptor Name	Publication Reference
GPR1	23 Genomics 609 (1994)
GPR4	14 DNA and Cell Biology 25 (1995)
GPR5	14 DNA and Cell Biology 25 (1995)
GPR7	28 Genomics 84 (1995)
GPR8	28 Genomics 84 (1995)
GPR9	184 J. Exp. Med. 963 (1996)
GPR10	29 Genomics 335 (1995)
GPR15	32 Genomics 462 (1996)
GPR17	70 J Neurochem. 1357 (1998)
GPR18	42 Genomics 462 (1997)
GPR20	187 Gene 75 (1997)
GPR21	187 Gene 75 (1997)
GPR22	187 Gene 75 (1997)
GPR24	398 FEBS Lett. 253 (1996)
GPR30	45 Genomics 607 (1997)
GPR31	42 Genomics 519 (1997)
GPR32	50 Genomics 281 (1997)
GPR40	239 Biochem. Biophys. Res. Commun. 543 (1997)
GPR41	239 Biochem. Biophys. Res. Commun. 543 (1997)
GPR43	239 Biochem. Biophys. Res. Commun. 543 (1997)
APJ	136 Gene 355 (1993)
BLR1	22 Eur. J. Immunol. 2759 (1992)
CEPR	231 Biochem. Biophys. Res. Commun. 651 (1997)
EBI1	23 Genomics 643 (1994)
EBI2	67 J. Virol. 2209 (1993)
ETBR-LP2	424 FEBS Lett. 193 (1998)
GPCR-CNS	54 Brain Res. Mol. Brain Res. 152 (1998); 45 Genomics 68 (1997)
GPR-NGA	394 FEBS Lett. 325 (1996)
H9	386 FEBS Lett 219 (1996)

- 7 -

HBA954	1261 Biochim. Biophys. Acta 121 (1995)
HG38	247 Biochem. Biophys. Res. Commun. 266 (1998)
HM74	5 Int. Immunol. 1239 (1993)
OGR1	35 Genomics 397 (1996)
V28	163 Gene 295 (1995)

5

As will be set forth and disclosed in greater detail below, utilization of a mutational cassette to modify the endogenous sequence of a human GPCR leads to a constitutively activated version of the human GPCR. These non-endogenous, constitutively activated versions of human GPCRs can be utilized, *inter alia*, for the screening of candidate compounds to directly identify compounds

10 of, *e.g.*, therapeutic relevance.

SUMMARY OF THE INVENTION

Disclosed herein is a non-endogenous, human G protein-coupled receptor comprising (a) as a most preferred amino acid sequence region (C-terminus to N-terminus orientation) and/or (b) as a most preferred nucleic acid sequence region (3' to 5' orientation) transversing

15 the transmembrane-6 (TM6) and intracellular loop-3 (IC3) regions of the GPCR:

(a) P¹ AA₁₅ X

wherein:

- 20
- (1) P¹ is an amino acid residue located within the TM6 region of the GPCR, where P¹ is selected from the group consisting of (i) the endogenous GPCR's proline residue, and (ii) a non-endogenous amino acid residue other than proline;
 - (2) AA₁₅ are 15 amino acids selected from the group consisting of

(a) the endogenous GPCR's amino acids (b) non-endogenous amino acid residues, and (c) a combination of the endogenous GPCR's amino acids and non-endogenous amino acids, excepting that none of the 15 endogenous amino acid residues that are positioned within the TM6 region of the GPCR is proline; and

- (3) X is a non-endogenous amino acid residue located within the IC3 region of said GPCR, preferably selected from the group consisting of lysine, histidine and arginine, and most preferably lysine, excepting that when the endogenous amino acid at position X is lysine, then X is an amino acid other than lysine, preferably alanine;

and/or

(b) P^{codon} (AA-codon)₁₅ X_{codon}

wherein:

- (1) P^{codon} is a nucleic acid sequence within the TM6 region of the GPCR, where P^{codon} encodes an amino acid selected from the group consisting of (i) the endogenous GPCR's proline residue, and (ii) a non-endogenous amino acid residue other than proline;
- (2) (AA-codon)₁₅ are 15 codons encoding 15 amino acids selected from the group consisting of (a) the endogenous GPCR's amino acids (b) non-endogenous amino acid residues and (c) a combination of the endogenous GPCR's amino acids and non-

- 9 -

endogenous amino acids, excepting that none of the 15 endogenous codons within the TM6 region of the GPCR encodes a proline amino acid residue; and

- (3) X_{codon} is a nucleic acid encoding region residue located within the IC3 region of said GPCR, where X_{codon} encodes a non-endogenous amino acid, preferably selected from the group consisting of lysine, histidine and arginine, and most preferably lysine, excepting that when the endogenous encoding region at position X_{codon} encodes the amino acid lysine, then X_{codon} encodes an amino acid other than lysine, preferably alanine.

The terms endogenous and non-endogenous in reference to these sequence cassettes are relative to the endogenous GPCR. For example, once the endogenous proline residue is located within the TM6 region of a particular GPCR, and the 16th amino acid therefrom is identified for mutation to constitutively activate the receptor, it is also possible to mutate the endogenous proline residue

(i.e., once the marker is located and the 16th amino acid to be mutated is identified, one may mutate the marker itself), although it is most preferred that the proline residue not be mutated. Similarly, and while it is most preferred that AA₁₅ be maintained in their endogenous forms, these amino acids may also be mutated. The only amino acid that must be mutated in the non-endogenous version of the human GPCR is X i.e., the endogenous amino acid that is 16 residues from P¹

cannot be maintained in its endogenous form and must be mutated, as further disclosed herein.

Stated again, while it is preferred that in the non-endogenous version of the human GPCR, P¹ and AA₁₅ remain in their endogenous forms (i.e., identical to their wild-type forms), once X is identified and mutated, any and/or all of P¹ and AA₁₅ can be mutated. This applies to the nucleic

acid sequences as well. In those cases where the endogenous amino acid at position X is lysine, then in the non-endogenous version of such GPCR, X is an amino acid other than lysine, preferably alanine.

Accordingly, and as a hypothetical example, if the endogenous GPCR has the following

5 endogenous amino acid sequence at the above-noted positions:

P-AACCTTGRRRDDDE -Q

then any of the following exemplary and hypothetical cassettes would fall within the scope of the disclosure (non-endogenous amino acids are set forth in bold):

P-AACCTTGRRRDDDE -K

10 P-AACCT**TH**IGRRRDDDE -K

P-ADEETTGGRRRDDDE -A

P-LLKFMSTWZLVAA**PQ** -K

A-LLKFMSTWZLVAA**PQ** -K

It is also possible to add amino acid residues within AA₁₅, but such an approach is not particularly
15 advanced. Indeed, in the most preferred embodiments, the only amino acid that differs in the non-endogenous version of the human GPCR as compared with the endogenous version of that GPCR is the amino acid in position X; mutation of this amino acid itself leads to constitutive activation of the receptor.

Thus, in particularly preferred embodiments, P¹ and P^{codon} are endogenous proline and an
20 endogenous nucleic acid encoding region encoding proline, respectively; and X and X_{codon} are non-endogenous lysine or alanine and a non-endogenous nucleic acid encoding region encoding lysine or alanine, respectively, with lysine being most preferred. Because it is most preferred that the non-endogenous versions of the human GPCRs which incorporate these mutations are

- 11 -

incorporated into mammalian cells and utilized for the screening of candidate compounds, the non-endogenous human GPCR incorporating the mutation need not be purified and isolated *per se* (i.e., these are incorporated within the cellular membrane of a mammalian cell), although such purified and isolated non-endogenous human GPCRs are well within the purview of this disclosure. Gene-
5 targeted and transgenic non-human mammals (preferably rats and mice) incorporating the non-endogenous human GPCRs are also within the purview of this invention; in particular, gene-targeted mammals are most preferred in that these animals will incorporate the non-endogenous versions of the human GPCRs in place of the non-human mammal's endogenous GPCR-encoding region (techniques for generating such non-human mammals to replace the non-human mammal's
10 protein encoding region with a human encoding region are well known; see, for example, U.S. Patent No. 5,777,194.)

It has been discovered that these changes to an endogenous human GPCR render the GPCR constitutively active such that, as will be further disclosed herein, the non-endogenous, constitutively activated version of the human GPCR can be utilized for, *inter alia*, the direct
15 screening of candidate compounds without the need for the endogenous ligand. Thus, methods for using these materials, and products identified by these methods are also within the purview of the following disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a generalized structure of a G protein-coupled receptor with the numbers
20 assigned to the transmembrane helices, the intracellular loops, and the extracellular loops.

Figure 2 schematically shows the two states, active and inactive, for a typical G protein coupled receptor and the linkage of the active state to the second messenger transduction pathway.

Figure 3 is a sequence diagram of the preferred vector pCMV, including restriction enzyme site locations.

Figure 4 is a diagrammatic representation of the signal measured comparing pCMV, non-endogenous, constitutively active GPR30 inhibition of GPR6-mediated activation of CRE-Luc reporter with endogenous GPR30 inhibition of GPR6-mediated activation of CRE-Luc reporter.

Figure 5 is a diagrammatic representation of the signal measured comparing pCMV, non-endogenous, constitutively activated GPR17 inhibition of GPR3-mediated activation of CRE-Luc reporter with endogenous GPR17 inhibition of GPR3-mediated activation of CRE-Luc reporter.

Figure 6 provides diagrammatic results of the signal measured comparing control pCMV, endogenous APJ and non-endogenous APJ.

Figure 7 provides an illustration of IP_3 production from non-endogenous human 5-HT_{2A} receptor as compared to the endogenous version of this receptor.

Figure 8 are dot-blot format results for GPR1 (8A), GPR30 (8B) and APJ (8C).

DETAILED DESCRIPTION

The scientific literature that has evolved around receptors has adopted a number of terms to refer to ligands having various effects on receptors. For clarity and consistency, the following definitions will be used throughout this patent document. To the extent that these definitions conflict with other definitions for these terms, the following definitions shall control:

AGONISTS shall mean compounds that activate the intracellular response when they bind to the receptor, or enhance GTP binding to membranes.

- 13 -

AMINO ACID ABBREVIATIONS used herein are set below:

	ALANINE	ALA	A
	ARGININE	ARG	R
	ASPARAGINE	ASN	N
5	ASPARTIC ACID	ASP	D
	CYSTEINE	CYS	C
	GLUTAMIC ACID	GLU	E
	GLUTAMINE	GLN	Q
	GLYCINE	GLY	G
10	HISTIDINE	HIS	H
	ISOLEUCINE	ILE	I
	LEUCINE	LEU	L
	LYSINE	LYS	K
	METHIONINE	MET	M
15	PHENYLALANINE	PHE	F
	PROLINE	PRO	P
	SERINE	SER	S
	THREONINE	THR	T
	TRYPTOPHAN	TRP	W
20	TYROSINE	TYR	Y
	VALINE	VAL	V

PARTIAL AGONISTS shall mean compounds which activate the intracellular response when they bind to the receptor to a lesser degree/extent than do agonists, or enhance GTP binding to membranes to a lesser degree/extent than do agonists

25 ANTAGONIST shall mean compounds that competitively bind to the receptor at the same site as the agonists but which do not activate the intracellular response initiated by the active form of the receptor, and can thereby inhibit the intracellular responses by agonists or partial agonists. ANTAGONISTS do not diminish the baseline intracellular response in the absence of an agonist or partial agonist.

30 CANDIDATE COMPOUND shall mean a molecule (for example, and not limitation, a chemical compound) which is amenable to a screening technique. Preferably, the phrase

"candidate compound" does not include compounds which were publicly known to be compounds selected from the group consisting of inverse agonist, agonist or antagonist to a receptor, as previously determined by an indirect identification process ("indirectly identified compound"); more preferably, not including an indirectly identified compound which has previously been
5 determined to have therapeutic efficacy in at least one mammal; and, most preferably, not including an indirectly identified compound which has previously been determined to have therapeutic utility in humans.

CODON shall mean a grouping of three nucleotides (or equivalents to nucleotides) which generally comprise a nucleoside (adenosine (A), guanosine (G), cytidine (C), uridine (U) and
10 thymidine (T)) coupled to a phosphate group and which, when translated, encodes an amino acid.

COMPOUND EFFICACY shall mean a measurement of the ability of a compound to inhibit or stimulate receptor functionality, as opposed to receptor binding affinity. A preferred means of detecting compound efficacy is via measurement of, *e.g.*, [³⁵S]GTPγS binding, as further disclosed in the Example section of this patent document.

15 **CONSTITUTIVELY ACTIVATED RECEPTOR** shall mean a receptor subject to constitutive receptor activation. In accordance with the invention disclosed herein, a non-endogenous, human constitutively activated G protein-coupled receptor is one that has been mutated to include the amino acid cassette P¹AA₁₅X, as set forth in greater detail below.

CONSTITUTIVE RECEPTOR ACTIVATION shall mean stabilization of a receptor
20 in the active state by means other than binding of the receptor with its endogenous ligand or a chemical equivalent thereof. Preferably, a G protein-coupled receptor subjected to constitutive receptor activation in accordance with the invention disclosed herein evidences at least a 10% difference in response (increase or decrease, as the case may be) to the signal measured for

- 15 -

constitutive activation as compared with the endogenous form of that GPCR, more preferably, about a 25% difference in such comparative response, and most preferably about a 50% difference in such comparative response. When used for the purposes of directly identifying candidate compounds, it is most preferred that the signal difference be at least about 50% such that there is
5 a sufficient difference between the endogenous signal and the non-endogenous signal to differentiate between selected candidate compounds. In most instances, the "difference" will be an increase in signal; however, with respect to Gs-coupled GPCRS, the "difference" measured is preferably a decrease, as will be set forth in greater detail below.

CONTACT or **CONTACTING** shall mean bringing at least two moieties together,
10 whether in an in vitro system or an in vivo system.

DIRECTLY IDENTIFYING or **DIRECTLY IDENTIFIED**, in relationship to the phrase "candidate compound", shall mean the screening of a candidate compound against a constitutively activated G protein-coupled receptor, and assessing the compound efficacy of such compound. This phrase is, under no circumstances, to be interpreted or understood to be
15 encompassed by or to encompass the phrase "indirectly identifying" or "indirectly identified."

ENDOGENOUS shall mean a material that is naturally produced by the genome of the species. **ENDOGENOUS** in reference to, for example and not limitation, GPCR, shall mean that which is naturally produced by a human, an insect, a plant, a bacterium, or a virus. By contrast, the term **NON-ENDOGENOUS** in this context shall mean that which is not naturally produced
20 by the genome of a species. For example, and not limitation, a receptor which is not constitutively active in its endogenous form, but when mutated by using the cassettes disclosed herein and thereafter becomes constitutively active, is most preferably referred to herein as a "non-endogenous, constitutively activated receptor." Both terms can be utilized to describe both "in

vivo" and "in vitro" systems. For example, and not limitation, in a screening approach, the endogenous or non-endogenous receptor may be in reference to an in vitro screening system whereby the receptor is expressed on the cell-surface of a mammalian cell. As a further example and not limitation, where the genome of a mammal has been manipulated to include a non-
5 endogenous constitutively activated receptor, screening of a candidate compound by means of an in vivo system is viable.

HOST CELL shall mean a cell capable of having a Plasmid and/or Vector incorporated therein. In the case of a prokaryotic Host Cell, a Plasmid is typically replicated as an autonomous molecule as the Host Cell replicates (generally, the Plasmid is thereafter isolated for introduction
10 into a eukaryotic Host Cell); in the case of a eukaryotic Host Cell, a Plasmid is integrated into the cellular DNA of the Host Cell such that when the eukaryotic Host Cell replicates, the Plasmid replicates. Preferably, for the purposes of the invention disclosed herein, the Host Cell is eukaryotic, more preferably, mammalian, and most preferably selected from the group consisting of 293, 293T and COS-7 cells.

15 **INDIRECTLY IDENTIFYING** or **INDIRECTLY IDENTIFIED** means the traditional approach to the drug discovery process involving identification of an endogenous ligand specific for an endogenous receptor, screening of candidate compounds against the receptor for determination of those which interfere and/or compete with the ligand-receptor interaction, and assessing the efficacy of the compound for affecting at least one second messenger pathway
20 associated with the activated receptor.

INHIBIT or **INHIBITING**, in relationship to the term "response" shall mean that a response is decreased or prevented in the presence of a compound as opposed to in the absence of the compound.

- 17 -

INVERSE AGONISTS shall mean compounds which bind to either the endogenous form of the receptor or to the constitutively activated form of the receptor, and which inhibit the baseline intracellular response initiated by the active form of the receptor below the normal base level of activity which is observed in the absence of agonists or partial agonists, or decrease GTP binding to membranes. Preferably, the baseline intracellular response is inhibited in the presence of the inverse agonist by at least 30%, more preferably by at least 50%, and most preferably by at least 75%, as compared with the baseline response in the absence of the inverse agonist.

KNOWN RECEPTOR shall mean an endogenous receptor for which the endogenous ligand specific for that receptor has been identified.

LIGAND shall mean an endogenous, naturally occurring molecule specific for an endogenous, naturally occurring receptor.

MUTANT or **MUTATION** in reference to an endogenous receptor's nucleic acid and/or amino acid sequence shall mean a specified change or changes to such endogenous sequences such that a mutated form of an endogenous, non-constitutively activated receptor evidences constitutive activation of the receptor. In terms of equivalents to specific sequences, a subsequent mutated form of a human receptor is considered to be equivalent to a first mutation of the human receptor if (a) the level of constitutive activation of the subsequent mutated form of the receptor is substantially the same as that evidenced by the first mutation of the receptor; and (b) the percent sequence (amino acid and/or nucleic acid) homology between the subsequent mutated form of the receptor and the first mutation of the receptor is at least about 80%, more preferably at least about 90% and most preferably at least 95%. Ideally, and owing to the fact that the most preferred cassettes disclosed herein for achieving constitutive activation includes a single amino acid and/or codon change between the endogenous and the non-endogenous forms of the GPCR (i.e. X or

X_{codon}), the percent sequence homology should be at least 98%.

ORPHAN RECEPTOR shall mean an endogenous receptor for which the endogenous ligand specific for that receptor has not been identified or is not known.

PHARMACEUTICAL COMPOSITION shall mean a composition comprising at least
5 one active ingredient, whereby the composition is amenable to investigation for a specified, efficacious outcome in a mammal (for example, and not limitation, a human). Those of ordinary skill in the art will understand and appreciate the techniques appropriate for determining whether an active ingredient has a desired efficacious outcome based upon the needs of the artisan.

PLASMID shall mean the combination of a Vector and cDNA. Generally, a Plasmid is
10 introduced into a Host Cell for the purpose of replication and/or expression of the cDNA as a protein.

STIMULATE or **STIMULATING**, in relationship to the term "response" shall mean that a response is increased in the presence of a compound as opposed to in the absence of the compound.

15 **TRANSVERSE** or **TRANSVERSING**, in reference to either a defined nucleic acid sequence or a defined amino acid sequence, shall mean that the sequence is located within at least two different and defined regions. For example, in an amino acid sequence that is 10 amino acid moieties in length, where 3 of the 10 moieties are in the TM6 region of a GPCR and the remaining
7 moieties are in the IC3 region of the GPCR, the 10 amino acid moiety can be described as
20 transverse the TM6 and IC3 regions of the GPCR.

VECTOR in reference to cDNA shall mean a circular DNA capable of incorporating at least one cDNA and capable of incorporation into a Host Cell.

The order of the following sections is set forth for presentational efficiency and is not

- 19 -

intended, nor should be construed, as a limitation on the disclosure or the claims to follow.

A. Introduction

The traditional study of receptors has always proceeded from the a priori assumption (historically based) that the endogenous ligand must first be identified before discovery could proceed to find antagonists and other molecules that could affect the receptor. Even in cases where an antagonist might have been known first, the search immediately extended to looking for the endogenous ligand. This mode of thinking has persisted in receptor research even after the discovery of constitutively activated receptors. What has not been heretofore recognized is that it is the active state of the receptor that is most useful for discovering agonists, partial agonists, and inverse agonists of the receptor. For those diseases which result from an overly active receptor or an under-active receptor, what is desired in a therapeutic drug is a compound which acts to diminish the active state of a receptor or enhance the activity of the receptor, respectively, not necessarily a drug which is an antagonist to the endogenous ligand. This is because a compound that reduces or enhances the activity of the active receptor state need not bind at the same site as the endogenous ligand. Thus, as taught by a method of this invention, any search for therapeutic compounds should start by screening compounds against the ligand-independent active state.

Screening candidate compounds against non-endogenous, constitutively activated GPCRs allows for the direct identification of candidate compounds which act at these cell surface receptors, without requiring any prior knowledge or use of the receptor's endogenous ligand. By determining areas within the body where the endogenous version of such GPCRs are expressed and/or over-expressed, it is possible to determine related disease/disorder states which are associated with the expression and/or over-expression of these receptors; such an approach is disclosed in this patent document.

B. Disease/Disorder Identification and/or Selection

Most preferably, inverse agonists to the non-endogenous, constitutively activated GPCRs can be identified using the materials of this invention. Such inverse agonists are ideal candidates as lead compounds in drug discovery programs for treating diseases related to these receptors.

- 5 Because of the ability to directly identify inverse agonists, partial agonists or agonists to these receptors, thereby allowing for the development of pharmaceutical compositions, a search, for diseases and disorders associated with these receptors is possible. For example, scanning both diseased and normal tissue samples for the presence of these receptor now becomes more than an academic exercise or one which might be pursued along the path of identifying, in the case of an
- 10 orphan receptor, an endogenous ligand. Tissue scans can be conducted across a broad range of healthy and diseased tissues. Such tissue scans provide a preferred first step in associating a specific receptor with a disease and/or disorder.

Preferably, the DNA sequence of the endogenous GPCR is used to make a probe for either radiolabeled cDNA or RT-PCR identification of the expression of the GPCR in tissue samples.

- 15 The presence of a receptor in a diseased tissue, or the presence of the receptor at elevated or decreased concentrations in diseased tissue compared to a normal tissue, can be preferably utilized to identify a correlation with that disease. Receptors can equally well be localized to regions of organs by this technique. Based on the known functions of the specific tissues to which the receptor is localized, the putative functional role of the receptor can be deduced.

20 **C. A "Human GPCR Proline Marker" Algorithm and the Creation of Non-Endogenous, Constitutively-Active Human GPCRs**

Among the many challenges facing the biotechnology arts is the unpredictability in gleaning genetic information from one species and correlating that information to another species

- 21 -

– nowhere in this art does this problem evidence more annoying exacerbation than in the genetic sequences that encode nucleic acids and proteins. Thus, for consistency and because of the highly unpredictable nature of this art, the following invention is limited, in terms of mammals, to human GPCRs – applicability of this invention to other mammalian species, while a potential possibility, is considered beyond mere rote application.

In general, when attempting to apply common "rules" from one related protein sequence to another or from one species to another, the art has typically resorted to sequence alignment, *i.e.*, sequences are linearized and attempts are then made to find regions of commonality between two or more sequences. While useful, this approach does not always prove to result in meaningful information. In the case of GPCRs, while the general structural motif is identical for all GPCRs, the variations in lengths of the TMs, ECs and ICs make such alignment approaches from one GPCR to another difficult at best. Thus, while it may be desirable to apply a consistent approach to, *e.g.*, constitutive activation from one GPCR to another, because of the great diversity in sequence length, fidelity, etc from one GPCR to the next, a generally applicable, and readily successful mutational alignment approach is in essence not possible. In an analogy, such an approach is akin to having a traveler start a journey at point A by giving the traveler dozens of different maps to point B, without any scale or distance markers on any of the maps, and then asking the traveler to find the shortest and most efficient route to destination B only by using the maps. In such a situation, the task can be readily simplified by having (a) a common "place-marker" on each map, and (b) the ability to measure the distance from the place-marker to destination B – this, then, will allow the traveler to select the most efficient from starting-point A to destination B.

In essence, a feature of the invention is to provide such coordinates within human GPCRs

that readily allows for creation of a constitutively active form of the human GPCRs.

As those in the art appreciate, the transmembrane region of a cell is highly hydrophobic; thus, using standard hydrophobicity plotting techniques, those in the art are readily able to determine the TM regions of a GPCR, and specifically TM6 (this same approach is also applicable to determining the EC and IC regions of the GPCR). It has been discovered that within the TM6 region of human GPCRs, a common proline residue (generally near the middle of TM6), acts as a constitutive activation "marker." By counting 15 amino acids from the proline marker, the 16th amino acid (which is located in the IC3 loop), when mutated from its endogenous form to a non-endogenous form, leads to constitutive activation of the receptor. For convenience, we refer to this as the "Human GPCR Proline Marker" Algorithm. Although the non-endogenous amino acid at this position can be any of the amino acids; most preferably, the non-endogenous amino acid is lysine. While not wishing to be bound by any theory, we believe that this position itself is unique and that the mutation at this location impacts the receptor to allow for constitutive activation.

We note that, for example, when the endogenous amino acid at the 16th position is already lysine (as is the case with GPR4 and GPR32), then in order for X to be a non-endogenous amino acid, it must be other than lysine; thus, in those situations where the endogenous GPCR has an endogenous lysine residue at the 16th position, the non-endogenous version of that GPCR preferably incorporates an amino acid other than lysine, preferably alanine, histidine and arginine, at this position. Of further note, it has been determined that GPR4 appears to be linked to Gs and active in its endogenous form (data not shown).

Because there are only 20 naturally occurring amino acids (although the use of non-naturally occurring amino acids is also viable), selection of a particular non-endogenous amino

- 23 -

acid for substitution at this 16th position is viable and allows for efficient selection of a non-endogenous amino acid that fits the needs of the investigator. However, as noted, the more preferred non-endogenous amino acids at the 16th position are lysine, histidine, arginine and alanine, with lysine being most preferred. Those of ordinary skill in the art are credited with the ability to readily determine proficient methods for changing the sequence of a codon to achieve a desired mutation.

It has also been discovered that occasionally, but not always, the proline residue marker will be preceded in TM6 by W2 (*i.e.*, W2P¹AA₁₅X) where W is tryptophan and 2 is any amino acid residue.

Our discovery, amongst other things, negates the need for unpredictable and complicated sequence alignment approaches commonly used by the art. Indeed, the strength of our discovery, while an algorithm in nature, is that it can be applied in a facile manner to human GPCRs, with dexterous simplicity by those in the art, to achieve a unique and highly useful end-product, *i.e.*, a constitutively activated version of a human GPCR. Because many years and significant amounts of money will be required to determine the endogenous ligands for the human GPCRs that the Human Genome project is uncovering, the disclosed invention not only reduces the time necessary to positively exploit this sequence information, but at significant cost-savings. This approach truly validates the importance of the Human Genome Project because it allows for the utilization of genetic information to not only understand the role of the GPCRs in, *e.g.*, diseases, but also provides the opportunity to improve the human condition.

D. Screening of Candidate Compounds

1. Generic GPCR screening assay techniques

When a G protein receptor becomes constitutively active, it couples to a G protein (*e.g.*,

Gq, Gs, Gi, Go) and stimulates release and subsequent binding of GTP to the G protein. The G protein then acts as a GTPase and slowly hydrolyzes the GTP to GDP, whereby the receptor, under normal conditions, becomes deactivated. However, constitutively activated receptors, including the non-endogenous, human constitutively active GPCRs of the present invention, continue to exchange GDP for GTP. A non-hydrolyzable analog of GTP, [³⁵S]GTP γ S, can be used to monitor enhanced binding to G proteins present on membranes which express constitutively activated receptors. It is reported that [³⁵S]GTP γ S can be used to monitor G protein coupling to membranes in the absence and presence of ligand. An example of this monitoring, among other examples well-known and available to those in the art, was reported by Traynor and Nahorski in 1995. The preferred use of this assay system is for initial screening of candidate compounds because the system is generically applicable to all G protein-coupled receptors regardless of the particular G protein that interacts with the intracellular domain of the receptor.

B 2. Specific GPCR screening assay techniques

C Once candidate compounds are identified using the "generic" G protein-coupled receptor assay (*i.e.*, an assay to select compounds that are agonists, partial agonists, or inverse agonists), further screening to confirm that the compounds have interacted at the receptor site is preferred. For example, a compound identified by the "generic" assay may not bind to the receptor, but may instead merely "uncouple" the G protein from the intracellular domain.

20 a. *Gs and Gi.*

Gs stimulates the enzyme adenylyl cyclase. Gi (and Go), on the other hand, inhibit this enzyme. Adenylyl cyclase catalyzes the conversion of ATP to cAMP; thus,

- 25 -

constitutively activated GPCRs that couple the Gs protein are associated with increased cellular levels of cAMP. On the other hand, constitutively activated GPCRs that couple the Gi (or Go) protein are associated with decreased cellular levels of cAMP. *See, generally, "Indirect Mechanisms of Synaptic Transmission," Chpt. 8, From Neuron To Brain (3rd Ed.)*

5 Nichols, J.G. et al eds. Sinauer Associates, Inc. (1992). Thus, assays that detect cAMP can be utilized to determine if a candidate compound is, *e.g.*, an inverse agonist to the receptor (*i.e.*, such a compound would decrease the levels of cAMP). A variety of approaches known in the art for measuring cAMP can be utilized; a most preferred approach relies upon the use of anti-cAMP antibodies in an ELISA-based format. Another type of assay that can be

10 utilized is a whole cell second messenger reporter system assay. Promoters on genes drive the expression of the proteins that a particular gene encodes. Cyclic AMP drives gene expression by promoting the binding of a cAMP-responsive DNA binding protein or transcription factor (CREB) which then binds to the promoter at specific sites called cAMP response elements and drives the expression of the gene. Reporter systems can be constructed which have a promoter containing

15 multiple cAMP response elements before the reporter gene, *e.g.*, β -galactosidase or luciferase. Thus, a constitutively activated Gs-linked receptor causes the accumulation of cAMP that then activates the gene and expression of the reporter protein. The reporter protein such as β -galactosidase or luciferase can then be detected using standard biochemical assays (Chen et al. 1995). With respect to GPCRs that link to Gi (or Go), and thus decrease levels of cAMP, an

20 approach to the screening of, *e.g.*, inverse agonists, based upon utilization of receptors that link to Gs (and thus increase levels of cAMP) is disclosed in the Example section with respect to GPR17 and GPR30.

b. Go and Gq.

Gq and Go are associated with activation of the enzyme phospholipase C, which in turn hydrolyzes the phospholipid PIP_2 , releasing two intracellular messengers: diacylglycerol (DAG) and inistol 1,4,5-triphoisphate (IP_3). Increased accumulation of IP_3 is associated with activation of Gq- and Go-associated receptors. *See, generally*, "Indirect Mechanisms of Synaptic Transmission," Chpt. 8, From Neuron To Brain (3rd Ed.) Nichols, J.G. et al eds. Sinauer Associates, Inc. (1992). Assays that detect IP_3 accumulation can be utilized to determine if a candidate compound is, *e.g.*, an inverse agonist to a Gq- or Go-associated receptor (*i.e.*, such a compound would decrease the levels of IP_3). Gq-associated receptors can also be examined using an AP1 reporter assay in that Gq-dependent phospholipase C causes activation of genes containing AP1 elements; thus, activated Gq-associated receptors will evidence an increase in the expression of such genes, whereby inverse agonists thereto will evidence a decrease in such expression, and agonists will evidence an increase in such expression. Commercially available assays for such detection are available.

E. Medicinal Chemistry

Generally, but not always, direct identification of candidate compounds is preferably conducted in conjunction with compounds generated via combinatorial chemistry techniques, whereby thousands of compounds are randomly prepared for such analysis. Generally, the results of such screening will be compounds having unique core structures; thereafter, these compounds are preferably subjected to additional chemical modification around a preferred core structure(s) to further enhance the medicinal properties thereof. Such techniques are

- 27 -

known to those in the art and will not be addressed in detail in this patent document.

F. Pharmaceutical Compositions

Candidate compounds selected for further development can be formulated into pharmaceutical compositions using techniques well known to those in the art. Suitable
5 pharmaceutically-acceptable carriers are available to those in the art; for example, *see* Remington's Pharmaceutical Sciences, 16th Edition, 1980, Mack Publishing Co., (Oslo et al., eds.)

G. Other Utility

Although a preferred use of the non-endogenous versions of the disclosed human GPCRs is for the direct identification of candidate compounds as inverse agonists, agonists or partial
10 agonists (preferably for use as pharmaceutical agents), these receptors can also be utilized in research settings. For example, in vitro and in vivo systems incorporating these receptors can be utilized to further elucidate and understand the roles of the receptors in the human condition, both normal and diseased, as well understanding the role of constitutive activation as it applies to understanding the signaling cascade. A value in these non-endogenous receptors is that their
15 utility as a research tool is enhanced in that, because of their unique features, the disclosed receptors can be used to understand the role of a particular receptor in the human body before the endogenous ligand therefor is identified. Other uses of the disclosed receptors will become apparent to those in the art based upon, *inter alia*, a review of this patent document.

EXAMPLES

20 The following examples are presented for purposes of elucidation, and not limitation, of the present invention. Following the teaching of this patent document that a mutational cassette may be utilized in the IC3 loop of human GPCRs based upon a position relative to a proline residue in TM6 to constitutively activate the receptor, and while specific nucleic acid

and amino acid sequences are disclosed herein, those of ordinary skill in the art are credited with the ability to make minor modifications to these sequences while achieving the same or substantially similar results reported below. Particular approaches to sequence mutations are within the purview of the artisan based upon the particular needs of the artisan.

5 Example 1

Preparation of Endogenous Human GPCRs

A variety of GPCRs were utilized in the Examples to follow. Some endogenous human GPCRs were graciously provided in expression vectors (as acknowledged below) and other endogenous human GPCRs were synthesized *de novo* using publicly-available sequence
10 information.

1. GPR1 (GenBank Accession Number: U13666)

The human cDNA sequence for GPR1 was provided in pRcCMV by Brian O'Dowd (University of Toronto). GPR1 cDNA (1.4kB fragment) was excised from the pRcCMV vector as a NdeI-XbaI fragment and was subcloned into the NdeI-XbaI site of pCMV vector (see
15 Figure 3). Nucleic acid (SEQ.ID.NO.: 1) and amino acid (SEQ.ID.NO.: 2) sequences for human GPR1 were thereafter determined and verified.

2. GPR4 (GenBank Accession Numbers: L36148, U35399, U21051)

The human cDNA sequence for GPR4 was provided in pRcCMV by Brian O'Dowd (University of Toronto). GPR1 cDNA (1.4kB fragment) was excised from the pRcCMV
20 vector as an ApaI(blunted)-XbaI fragment and was subcloned (with most of the 5' untranslated region removed) into HindIII(blunted)-XbaI site of pCMV vector. Nucleic acid (SEQ.ID.NO.: 3) and amino acid (SEQ.ID.NO.: 4) sequences for human GPR4 were thereafter determined and verified.

3. GPR5 (GenBank Accession Number: L36149)

- 29 -

The cDNA for human GPR5 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 64°C for 1min; and 72°C for 1.5 min. The 5' PCR primer contained an EcoRI site with the sequence: 5'-TATGAATTCAGATGCTCTAAACGTCCCTGC-3' (SEQ.ID.NO.: 5) and the 3' primer contained BamHI site with the sequence: 5'-TCCGGATCCACCTGCACCTGCGCCTGCACC-3' (SEQ.ID.NO.: 6). The 1.1 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 7) and amino acid (SEQ.ID.NO.: 8) sequences for human GPR5 were thereafter determined and verified.

4. GPR7 (GenBank Accession Number: U22491)

The cDNA for human GPR7 was generated and cloned into pCMV expression vector as follows: PCR condition- PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 62°C for 1min; and 72°C for 1min and 20 sec. The 5' PCR primer contained a HindIII site with the sequence:

5'-GCAAGCTTGGGGGACGCCAGGTCGCCGGCT-3' (SEQ.ID.NO.: 9)

and the 3' primer contained a BamHI site with the sequence:

5'-GCGGATCCGGACGCTGGGGGAGTCAGGCTGC-3' (SEQ.ID.NO.: 10).

The 1.1 kb PCR fragment was digested with HindIII and BamHI and cloned into HindIII-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 11) and amino acid (SEQ.ID.NO.:

12) sequences for human GPR7 were thereafter determined and verified.

5. **GPR8 (GenBank Accession Number: U22492)**

The cDNA for human GPR8 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase

5 (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 62°C for 1min; and 72 °C for 1min and 20 sec. The 5' PCR primer contained an EcoRI site with the sequence:

5'-CGGAATTCGTCAACGGTCCCAGCTACAATG-3' (SEQ.ID.NO.: 13).

10 and the 3' primer contained a BamHI site with the sequence: .

5'-ATGGATCCCAGGCCCTTCAGCACCGCAATAT-3'(SEQ.ID.NO.: 14).

The 1.1 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of PCMV expression vector. All 4 cDNA clones sequenced contained a possible polymorphism involving a change of amino acid 206 from Arg to Gln. Aside from this

15 difference, nucleic acid (SEQ.ID.NO.: 15) and amino acid (SEQ.ID.NO.: 16) sequences for human
GPR8 were thereafter determined and verified.

6. **GPR9 (GenBank Accession Number: X95876)**

The cDNA for human GPR9 was generated and cloned into pCMV expression vector as follows: PCR was performed using a clone (provided by Brian O'Dowd) as template and

20 pfu polymerase (Stratagene) with the buffer system provided by the manufacturer supplemented with 10% DMSO, 0.25 μ M of each primer, and 0.5 mM of each of the 4 nucleotides. The cycle condition was 25 cycles of: 94°C for 1 min; 56°C for 1min; and 72°C for 2.5 min. The 5' PCR primer contained an EcoRI site with the sequence:

- 31 -

5'-ACGAATTCAGCCATGGTCCTTGAGGTGAGTGACCACCAAGTGCTAAAT-3'
(SEQ.ID.NO.: 17)

and the 3' primer contained a BamHI site with the sequence:

5'-GAGGATCCTGGAATGCGGGGAAGTCAG-3' (SEQ.ID.NO.: 18).

- 5 The 1.2 kb PCR fragment was digested with EcoRI and cloned into EcoRI-SmaI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 19) and amino acid (SEQ.ID.NO.: 20) sequences for human GPR9 were thereafter determined and verified.

7. GPR9-6 (GenBank Accession Number: U45982)

The cDNA for human GPR9-6 was generated and cloned into pCMV expression
10 vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 62°C for 1min; and 72°C for 1 min and 20 sec. The 5' PCR primer was kinased with the sequence:
5'-TTAAGCTTGACCTAATGCCATCTTGTGTCC-3' (SEQ.ID.NO.: 21)

- 15 and the 3' primer contained a BamHI site with the sequence:

5'-TTGGATCCAAAAGAACCATGCACCTCAGAG-3' (SEQ.ID.NO.: 22).

The 1.2 kb PCR fragment was digested with BamHI and cloned into EcoRV-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 23) and amino acid (SEQ.ID.NO.: 24) sequences for human GPR9-6 were thereafter determined and verified.

20 8. GPR10 (GenBank Accession Number: U32672)

The human cDNA sequence for GPR10 was provided in pRcCMV by Brian O'Dowd (University of Toronto). GPR10 cDNA (1.3kB fragment) was excised from the pRcCMV vector as an EcoRI-XbaI fragment and was subcloned into EcoRI-XbaI site of pCMV

vector. Nucleic acid (SEQ.ID.NO.: 25) and amino acid (SEQ.ID.NO.: 26) sequences for human GPR10 were thereafter determined and verified.

9. GPR15 (GenBank Accession Number: U34806)

The human cDNA sequence for GPR15 was provided in pCDNA3 by Brian O'Dowd (University of Toronto). GPR15 cDNA (1.5kB fragment) was excised from the pCDNA3 vector as a HindIII-Bam fragment and was subcloned into HindIII-Bam site of pCMV vector. Nucleic acid (SEQ.ID.NO.: 27) and amino acid (SEQ.ID.NO.: 28) sequences for human GPR15 were thereafter determined and verified.

10. GPR17 (GenBank Accession Number: Z94154)

The cDNA for human GPR17 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 56°C for 1min and 72 °C for 1 min and 20 sec. The 5' PCR primer contained an EcoRI site with the sequence:

5'-CTAGAATTCTGACTCCAGCCAAAGCATGAAT-3' (SEQ.ID.NO.: 29) and the 3' primer contained a BamHI site with the sequence:

5'-GCTGGATCCTAAACAGTCTGCGCTCGGCCT-3' (SEQ.ID.NO.: 30).

The 1.1 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 31) and amino acid (SEQ.ID.NO.: 32) sequences for human GPR17 were thereafter determined and verified.

11. GPR18 (GenBank Accession Number: L42324)

The cDNA for human GPR18 was generated and cloned into pCMV expression

- 33 -

vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 54°C for 1min; and 72 °C for 1min and 20 sec. The 5' PCR primer was kinased with the sequence:

5 5'-ATAAGATGATCACCCTGAACAATCAAGAT -3' (SEQ.ID.NO.: 33)

and the 3' primer contained an EcoRI site with the sequence:

5'-TCCGAATTCATAACATTTCACCTGTTTATATTGC-3' (SEQ.ID.NO.: 34).

The 1.0 kb PCR fragment was digested with EcoRI and cloned into blunt-EcoRI site of pCMV expression vector. All 8 cDNA clones sequenced contained 4 possible polymorphisms involving
10 changes of amino acid 12 from Thr to Pro, amino acid 86 from Ala to Glu, amino acid 97 from Ile to Leu and amino acid 310 from Leu to Met. Aside from these changes, nucleic acid (SEQ.ID.NO.: 35) and amino acid (SEQ.ID.NO.: 36) sequences for human GPR18 were thereafter determined and verified.

12. GPR20 (GenBank Accession Number: U66579)

15 The cDNA for human GPR20 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 62°C for 1min; and 72 °C for 1 min and 20 sec. The 5' PCR primer was kinased with the sequence:

20 5'-CCAAGCTTCCAGGCCTGGGGTGTGCTGG-3' (SEQ.ID.NO.: 37)

and the 3' primer contained a BamHI site with the sequence:

5'-ATGGATCCTGACCTTCGGCCCCCTGGCAGA-3' (SEQ.ID.NO.: 38).

The 1.2 kb PCR fragment was digested with BamHI and cloned into EcoRV-BamHI site of

PCMV expression vector. Nucleic acid (SEQ.ID.NO.: 39) and amino acid (SEQ.ID.NO.: 40) sequences for human GPR20 were thereafter determined and verified.

13. GPR21 (GenBank Accession Number: U66580)

The cDNA for human GPR21 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 62°C for 1min; and 72°C for 1 min and 20 sec. The 5' PCR primer was kinased with the sequence:

5'-GAGAATTCACCTCCTGAGCTCAAGATGAACT-3' (SEQ.ID.NO.: 41)

and the 3' primer contained a BamHI site with the sequence:

5'-CGGGATCCCCGTAAGTCTGAGCCACTTCAGAT-3' (SEQ.ID.NO.: 42).

The 1.1 kb PCR fragment was digested with BamHI and cloned into EcoRV-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 43) and amino acid (SEQ.ID.NO.: 44) sequences for human GPR21 were thereafter determined and verified.

14. GPR22 (GenBank Accession Number: U66581)

The cDNA for human GPR22 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 50°C for 1min; and 72°C for 1.5 min. The 5' PCR primer was kinased with the sequence:

5'-TCCCCCGGGAAAAAACCAGTCTCCAAA-3' (SEQ.ID.NO.: 45)

and the 3' primer contained a BamHI site with the sequence:

5'-TAGGATCCATTTGAATGTGGATTTGGTGAAA-3' (SEQ.ID.NO.: 46).

- 35 -

The 1.38 kb PCR fragment was digested with BamHI and cloned into EcoRV-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 47) and amino acid (SEQ.ID.NO.: 48) sequences for human GPR22 were thereafter determined and verified.

15. GPR24 (GenBank Accession Number: U71092)

5 The cDNA for human GPR24 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 56°C for 1min; and 72 °C for 1 min and 20 sec. The 5' PCR primer contains a HindIII site with the
10 sequence:

5'-GTGAAGCTTGCCTCTGGTGCCTGCAGGAGG-3' (SEQ.ID.NO.: 49)

and the 3' primer contains an EcoRI site with the sequence:

5'-GCAGAATTCCCGGTGGCGTGTTGTGGTGCCC-3' (SEQ.ID.NO.: 50).

The 1.3 kb PCR fragment was digested with HindIII and EcoRI and cloned into HindIII-EcoRI
15 site of pCMV expression vector. The nucleic acid (SEQ.ID.NO.: 51) and amino acid sequence (SEQ.ID.NO.: 52) for human GPR24 were thereafter determined and verified.

16. GPR30 (GenBank Accession Number: U63917)

The cDNA for human GPR30 was generated and cloned as follows: the coding sequence of GPR30 (1128bp in length) was amplified from genomic DNA using the primers:
20 5'-GGCGGATCCATGGATGTGACTTCCCAA-3' (SEQ.ID.NO.: 53) and
5'-GGCGGATCCCTACACGGCACTGCTGAA-3' (SEQ.ID.NO.: 54).

The amplified product was then cloned into a commercially available vector, pCR2.1 (Invitrogen),

using a "TOPO-TA Cloning Kit" (Invitrogen, #K4500-01), following manufacturer instructions.

The full-length GPR30 insert was liberated by digestion with BamHI, separated from the vector by agarose gel electrophoresis, and purified using a Sephaglas Bandprep™ Kit (Pharmacia, # 27-9285-01) following manufacturer instructions. The nucleic acid (SEQ.ID.NO.: 55) and amino acid

5 sequence (SEQ.ID.NO.: 56) for human GPR30 were thereafter determined and verified.

17. GPR31 (GenBank Accession Number: U65402)

The cDNA for human GPR31 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and

10 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 58°C for 1min; and 72 °C for 2 min. The 5' PCR primer contained an EcoRI site with the sequence:

5'-AAGGAATTCACGGCCGGGTGATGCCATTCCC-3' (SEQ.ID.NO.: 57)

and the 3' primer contained a BamHI site with the sequence:

5'-GGTGGATCCATAAACACGGGCGTTGAGGAC -3' (SEQ.ID.NO.: 58).

15 The 1.0 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 59) and amino acid (SEQ.ID.NO.: 60) sequences for human GPR31 were thereafter determined and verified.

18. GPR32 (GenBank Accession Number: AF045764)

The cDNA for human GPR32 was generated and cloned into pCMV expression
20 vector as follows: PCR was performed using genomic DNA as template and rTth polymerase

(Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 56°C for 1min; and 72 °C for 1 min and 20 sec. The 5' PCR primer contained an EcoRI site with the

- 37 -

sequence:

5'-TAAGAATTCCATAAAAAATTATGGAATGG-3' (SEQ.ID.NO.:243)

and the 3' primer contained a BamHI site with the sequence:

5'-CCAGGATCCAGCTGAAGTCTTCCATCATTC-3' (SEQ.ID.NO.: 244).

- 5 The 1.1 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 245) and amino acid (SEQ.ID.NO.: 246) sequences for human GPR32 were thereafter determined and verified.

19. GPR40 (GenBank Accession Number: AF024687)

- The cDNA for human GPR40 was generated and cloned into pCMV expression
10 vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min, 65°C for 1min and 72 °C for 1 min and 10 sec. The 5' PCR primer contained an EcoRI site with the sequence

- 15 5'-GCAGAATTCGGCGGCCCCATGGACCTGCCCCC-3' (SEQ.ID.NO.: 247)

and the 3' primer contained a BamHI site with the sequence

5'-GCTGGATCCCCCGAGCAGTGGCGTTACTTC-3' (SEQ.ID.NO.: 248).

- The 1 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 249) and amino acid (SEQ.ID.NO.: 250)
20 sequences for human GPR40 were thereafter determined and verified.

20. GPR41 (GenBank Accession Number AF024688)

The cDNA for human GPR41 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase

(Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of 94°C for 1 min, 65°C for 1min and 72 °C for 1 min and 10 sec. The 5' PCR primer contained an HindIII site with the sequence:

5 5'-CTCAAGCTTACTCTCTCTCACCAGTGGCCAC-3' (SEQ.ID.NO.: 251)

and the 3' primer was kinased with the sequence

5'-CCCTCCTCCCCCGGAGGACCTAGC-3' (SEQ.ID.NO.: 252).

The 1 kb PCR fragment was digested with HindIII and cloned into HindIII-blunt site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 253) and amino acid (SEQ.ID.NO.: 254)

10 sequences for human GPR41 were thereafter determined and verified.

21. GPR43 (GenBank Accession Number AF024690)

The cDNA for human GPR43 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and

15 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1min; and 72 °C for 1 min and 10 sec. The 5' PCR primer contains an HindIII site with the sequence:

5'-TTTAAGCTTCCCCTCCAGGATGCTGCCGGAC-3' (SEQ.ID.NO.: 255)

and the 3' primer contained an EcoRI site with the sequence:

20 5'-GGCGAATTCTGAAGGTCCAGGGAACTGCTA-3' (SEQ.ID.NO. 256).

The 1 kb PCR fragment was digested with HindIII and EcoRI and cloned into HindIII-EcoRI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 257) and amino acid (SEQ.ID.NO.: 258) sequences for human GPR43 were thereafter determined and verified.

- 39 -

22. APJ (GenBank Accession Number: U03642)

Human APJ cDNA (in pRcCMV vector) was provided by Brian O'Dowd (University of Toronto). The human APJ cDNA was excised from the pRcCMV vector as an EcoRI-XbaI (blunted) fragment and was subcloned into EcoRI-SmaI site of pCMV vector.

- 5 Nucleic acid (SEQ.ID.NO.: 61) and amino acid (SEQ.ID.NO.: 62) sequences for human APJ were thereafter determined and verified.

23. BLR1 (GenBank Accession Number: X68149)

The cDNA for human BLR1 was generated and cloned into pCMV expression vector as follows: PCR was performed using thymus cDNA as template and rTth polymerase
10 (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 62°C for 1min; and 72 °C for 1 min and 20 sec. The 5' PCR primer contained an EcoRI site with the sequence:

5'-TGAGAATTCTGGTGACTCACAGCCGGCACAG-3' (SEQ.ID.NO.: 63):

- 15 and the 3' primer contained a BamHI site with the sequence:

5'-GCCGGATCCAAGGAAAAGCAGCAATAAAAGG-3' (SEQ.ID.NO.: 64). The 1.2 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 65) and amino acid (SEQ.ID.NO.: 66) sequences for human BLR1 were thereafter determined and verified.

20 **24. CEPR (GenBank Accession Number: U77827)**

The cDNA for human CEPR was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and

0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1min; and 72 °C for 1 min and 20 sec. The 5' PCR primer was kinased with the sequence:

5'-CAAAGCTTGAAAGCTGCACGGTGCAGAGAC-3' (SEQ.ID.NO.:67)

and the 3' primer contained a BamHI site with the sequence:

5 5'-GCGGATCCCGAGTCACACCCTGGCTGGGCC-3' (SEQ.ID.NO.: 68).

The 1.2 kb PCR fragment was digested with BamHI and cloned into EcoRV-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 69) and amino acid (SEQ.ID.NO.: 70) sequences for human CEPR were thereafter determined and verified.

25. EBI1 (GenBank Accession Number: L31581)

10 The cDNA for human EBI1 was generated and cloned into pCMV expression vector as follows: PCR was performed using thymus cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 62°C for 1min; and 72 °C for 1 min and 20 sec. The 5' PCR primer contained an EcoRI site with the
15 sequence:

5'-ACAGAATTCCTGTGTGGTTTTACCGCCCAG-3' (SEQ.ID.NO.: 71)

and the 3' primer contained a BamHI site with the sequence:

5'-CTCGGATCCAGGCAGAAGAGTCGCCTATGG-3' (SEQ.ID.NO.: 72).

The 1.2 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI
20 site of PCMV expression vector. Nucleic acid (SEQ.ID.NO.: 73) and amino acid (SEQ.ID.NO.: 74) sequences for human EBI1 were thereafter determined and verified.

26. EBI2 (GenBank Accession Number: L08177)

The cDNA for human EBI2 was generated and cloned into pCMV expression

- 41 -

vector as follows: PCR was performed using cDNA clone (graciously provided by Kevin Lynch, University of Virginia Health Sciences Center; the vector utilized was not identified by the source) as template and pfu polymerase (Stratagene) with the buffer system provided by the manufacturer supplemented with 10% DMSO, 0.25 μ M of each primer, and 0.5 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 60°C for 1min; and 72°C for 1 min and 20 sec. The 5' PCR primer contained an EcoRI site with the sequence:

5'-CTGGAATTCACCTGGACCACCACCAATGGATA-3' (SEQ.ID.NO.: 75)

and the 3' primer contained a BamHI site with the sequence

5'-CTCGGATCCTGCAAAGTTTGTACATACAG TT-3' (SEQ.ID.NO.: 76).

- 10 The 1.2 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 77) and amino acid (SEQ.ID.NO.: 78) sequences for human EBI2 were thereafter determined and verified.

27. ETBR-LP2 (GenBank Accession Number: D38449)

The cDNA for human ETBR-LP2 was generated and cloned into pCMV expression vector as follows: PCR was performed using brain cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1min; and 72 °C for 1.5 min. The 5' PCR contained an EcoRI site with the sequence:

20 5'-CTGGAATTCTCCTGCTCATCCAGCCATGCGG -3' (SEQ.ID.NO.: 79)

and the 3' primer contained a BamHI site with the sequence:

5'-CCTGGATCCCCACCCCTACTGGGGCCTCAG -3' (SEQ.ID.NO.: 80).

The 1.5 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI

site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 81) and amino acid (SEQ.ID.NO.: 82) sequences for human ETBR-LP2 were thereafter determined and verified.

28. GHSR (GenBank Accession Number: U60179)

The cDNA for human GHSR was generated and cloned into pCMV expression
5 vector as follows: PCR was performed using hippocampus cDNA as template and TaqPlus Precision polymerase (Stratagene) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 68°C for 1min; and 72°C for 1 min and 10 sec. For first round PCR, the 5' PCR primer sequence was:

10 5'-ATGTGGAACGCGACGCCAGCG-3' (SEQ.ID.NO.: 83)

and the 3' primer sequence was:

5'-TCATGTATTAATACTAGATTCT-3' (SEQ.ID.NO.: 84).

Two microliters of the first round PCR was used as template for the second round PCR where the 5' primer was kinased with sequence:

15 5'-TACCATGTGGAACGCGACGCCAGCGAAGAGCCGGGGT-3' (SEQ.ID.NO.:85)

and the 3' primer contained an EcoRI site with the sequence:

5'-CGGAATTCATGTATTAATACTAGATTCTGTCCAGGCCCG-3' (SEQ.ID.NO.:86).

The 1.1 kb PCR fragment was digested with EcoRI and cloned into blunt-EcoRI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 87) and amino acid (SEQ.ID.NO.: 88) sequences
20 for human GHSR were thereafter determined and verified.

29. GPCR-CNS (GenBank Accession Number: AF017262)

The cDNA for human GPCR-CNS was generated and cloned into pCMV expression vector as follows: PCR was performed using brain cDNA as template and rTth

- 43 -

polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1min; and 72°C for 2 min. The 5' PCR primer contained a HindIII site with the sequence:

5 5'-GCAAGCTTGTGCCCTCACCAAGCCATGCGAGCC-3' (SEQ.ID.NO.: 89)

and the 3' primer contained an EcoRI site with the sequence:

5'-CGGAATTCAGCAATGAGTTCCGACAGAAGC-3' (SEQ.ID.NO.: 90).

The 1.9 kb PCR fragment was digested with HindIII and EcoRI and cloned into HindIII-EcoRI site of pCMV expression vector. All nine clones sequenced contained a potential polymorphism
10 involving a S284C change. Aside from this difference, nucleic acid (SEQ.ID.NO.: 91) and amino acid (SEQ.ID.NO.: 92) sequences for human GPCR-CNS were thereafter determined and verified.

30. GPR-NGA (GenBank Accession Number: U55312)

The cDNA for human GPR-NGA was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth
15 polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of 94°C for 1 min, 56°C for 1min and 72°C for 1.5 min. The 5' PCR primer contained an EcoRI site with the sequence:

5'-CAGAATTCAGAGAAAAAAGTGAATATGGTTTTT-3' (SEQ.ID.NO.: 93)

20 and the 3' primer contained a BamHI site with the sequence:

5'-TTGGATCCCTGGTGCATAACAATTGAAAGAAT-3' (SEQ.ID.NO.: 94).

The 1.3 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 95) and amino acid (SEQ.ID.NO.:

96) sequences for human GPR-NGA were thereafter determined and verified.

31. H9 (GenBank Accession Number: U52219)

The cDNA for human HB954 was generated and cloned into pCMV expression vector as follows: PCR was performed using pituitary cDNA as template and rTth polymerase

5 (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min, 62°C for 1min and 72 °C for 2 min. The 5' PCR primer contains a HindIII site with the sequence:

5'-GGAAAGCTTAACGATCCCCAGGAGCAACAT-3' (SEQ.ID.NO.: 97)

and the 3' primer contains a BamHI site with the sequence:

10 5'-CTGGGATCCTACGAGAGCATTTTTCACACAG-3' (SEQ.ID.NO.: 98).

The 1.9 kb PCR fragment was digested with HindIII and BamHI and cloned into HindIII-BamHI site of pCMV expression vector. When compared to the published sequences, a different isoform with 12 bp in frame insertion in the cytoplasmic tail was also identified and designated "H9b." Both isoforms contain two potential polymorphisms involving changes

15 of amino acid P320S and amino acid G448A. Isoform H9a contained another potential polymorphism of amino acid S493N, while isoform H9b contained two additional potential polymorphisms involving changes of amino acid I502T and amino acid A532T (corresponding to amino acid 528 of isoform H9a). Nucleic acid (SEQ.ID.NO.: 99) and amino acid (SEQ.ID.NO.: 100) sequences for human H9 were thereafter determined and
20 verified (in the section below, both isoforms were mutated in accordance with the Human GPCR Proline Marker Algorithm).

32. HB954 (GenBank Accession Number: D38449)

The cDNA for human HB954 was generated and cloned into pCMV expression

- 45 -

vector as follows: PCR was performed using brain cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of 94°C for 1 min, 58°C for 1min and 72°C for 2 min. The 5' PCR contained a HindIII site with the sequence:

5 5'-TCCAAGCTTCGCCATGGGACATAACGGGAGCT -3' (SEQ.ID.NO.: 101)

and the 3' primer contained an EcoRI site with the sequence:

5'-CGTGAATTCCAAGAATTTACAATCCTTGCT -3' (SEQ.ID.NO.: 102).

The 1.6 kb PCR fragment was digested with HindIII and EcoRI and cloned into HindIII-EcoRI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 103) and amino acid
10 (SEQ.ID.NO.: 104) sequences for human HB954 were thereafter determined and verified:

33. HG38 (GenBank Accession Number: AF062006)

The cDNA for human HG38 was generated and cloned into pCMV expression vector as follows: PCR was performed using brain cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM
15 of each 4 nucleotides. The cycle condition was 30 cycles of 94°C for 1 min, 56°C for 1min and 72°C for 1 min and 30 sec. Two PCR reactions were performed to separately obtain the 5' and 3' fragment. For the 5' fragment, the 5' PCR primer contained an HindIII site with the sequence:

5'-CCCAAGCTTCGGGCACCATGGACACCTCCC-3' (SEQ.ID.NO.: 259)

and the 3' primer contained a BamHIsite with the sequence:

20 5'-ACAGGATCCAAATGCACAGCACTGGTAAGC-3' (SEQ.ID.NO.: 260).

This 5' 1.5 kb PCR fragment was digested with HindIII and BamHI and cloned into an HindIII-BamHI site of pCMV. For the 3' fragment, the 5' PCR primer was kinased with the sequence:

5'-CTATAACTGGGTACATGGTTTAAC-3' (SEQ.ID.NO. 261)

and the 3' primer contained an EcoRI site with the sequence:

5'-TTTGAATTCACATATTAATTAGAGACATGG-3' (SEQ.ID.NO.: 262).

The 1.4 kb 3' PCR fragment was digested with EcoRI and subcloned into a blunt-EcoRI site of pCMV vector. The 5' and 3' fragments were then ligated together through a common EcoRV site
5 to generate the full length cDNA clone. Nucleic acid (SEQ.ID.NO.: 263) and amino acid (SEQ.ID.NO.: 264) sequences for human HG38 were thereafter determined and verified.

34. HM74 (GenBank Accession Number: D10923)

The cDNA for human HM74 was generated and cloned into pCMV expression vector as follows: PCR was performed using either genomic DNA or thymus cDNA (pooled) as
10 template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1min; and 72°C for 1 min and 20 sec. The
5' PCR primer contained an EcoRI site with the sequence:

5'-GGAGAATTCAGTAGGCGAGGCGCTCCATC-3' (SEQ.ID.NO.: 105)

15 and the 3' primer was kinased with the sequence:

5'-GGAGGATCCAGGAAACCTTAGGCCGAGTCC-3' (SEQ.ID.NO.:106).

The 1.3 kb PCR fragment was digested with EcoRI and cloned into EcoRI-SmaI site of pCMV expression vector. Clones sequenced revealed a potential polymorphism involving a
N94K change. Aside from this difference, nucleic acid (SEQ.ID.NO.: 107) and amino acid
20 (SEQ.ID.NO.: 108) sequences for human HM74 were thereafter determined and verified.

35. MIG (GenBank Accession Numbers: AFO44600 and AFO44601)

The cDNA for human MIG was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and TaqPlus Precision

- 47 -

polymerase (Stratagene) for first round PCR or pfu polymerase (Stratagene) for second round PCR with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM (TaqPlus Precision) or 0.5 mM (pfu) of each of the 4 nucleotides. When pfu was used, 10% DMSO was included in the buffer. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1 min; and 72 °C for: (a) 1 min for first round PCR; and (b) 2 min for second round PCR.

Because there is an intron in the coding region, two sets of primers were separately used to generate overlapping 5' and 3' fragments. The 5' fragment PCR primers were:

5'-ACCATGGCTTGCAATGGCAGTGCGGCCAGGGGGGCACT-3' (external sense)
(SEQ.ID.NO.: 109)

10 and

5'-CGACCAGGACAAACAGCATCTTGGTCACTTGTCTCCGGC-3' (internal antisense)
(SEQ.ID.NO.: 110).

The 3' fragment PCR primers were:

5'-GACCAAGATGCTGTTTGTCTCCTGGTCGTGGTGTGTTGGCAT-3' (internal sense)
15 (SEQ.ID.NO.: 111) and

5'-CGGAATTCAGGATGGATCGGTCTCTTGCTGCGCCT-3' (external antisense with an
EcoRI site) (SEQ.ID.NO.: 112).

The 5' and 3' fragments were ligated together by using the first round PCR as template and the kinased external sense primer and external antisense primer to perform second round PCR. The
20 1.2 kb PCR fragment was digested with EcoRI and cloned into the blunt-EcoRI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 113) and amino acid (SEQ.ID.NO.: 114) sequences for human MIG were thereafter determined and verified.

36. OGR1 (GenBank Accession Number: U48405)

The cDNA for human OGR1 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1min; and 72°C for 1 min and 20 sec. The 5' PCR primer was kinased with the sequence: 5'-GGAAGCTTCAGGCCCAAAGATGGGGAACAT-3' (SEQ.ID.NO.: 115); and the 3' primer contained a BamHI site with the sequence: 5'-GTGGATCCACCCGCGGAGGACCCAGGCTAG -3' (SEQ.ID.NO.: 116). The 1.1 kb PCR fragment was digested with BamHI and cloned into the EcoRV-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 117) and amino acid (SEQ.ID.NO.: 118) sequences for human OGR1 were thereafter determined and verified.

37. Serotonin 5HT_{2A}

The cDNA encoding endogenous human 5HT_{2A} receptor was obtained by RT-PCR using human brain poly-A⁺ RNA; a 5' primer from the 5' untranslated region with an Xho I restriction site:

5'-GACCTCGAGTCCTTCTACACCTCATC-3' (SEQ.ID.NO: 119)

and a 3' primer from the 3' untranslated region containing an Xba I site:

5'-TGCTCTAGATTCCAGATAGGTGAAAACCTTG-3' (SEQ.ID.NO: 120)

PCR was performed using either TaqPlus™ precision polymerase (Stratagene) or rTth™ polymerase (Perkin Elmer) with the buffer system provided by the manufacturers, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 57°C for 1min; and 72°C for 2 min. The 1.5 kb PCR fragment was digested with Xba I and subcloned into Eco RV-Xba I site of pBluescript. The resulting cDNA clones were fully

- 49 -

sequenced and found to encode two amino acid changes from the published sequences. The first one was a T25N mutation in the N-terminal extracellular domain; the second is an H452Y mutation. Because cDNA clones derived from two independent PCR reactions using Taq polymerase from two different commercial sources (TaqPlus™ from Stratagene and rTth™ Perkin Elmer) contained the same two mutations, these mutations are likely to represent sequence polymorphisms rather than PCR errors. With these exceptions, the nucleic acid (SEQ.ID.NO.: 121) and amino acid (SEQ.ID.NO.: 122) sequences for human 5HT_{2A} were thereafter determined and verified.

38. Serotonin 5HT_{2C}

The cDNA encoding endogenous human 5HT_{2C} receptor was obtained from human brain poly-A⁺ RNA by RT-PCR. The 5' and 3' primers were derived from the 5' and 3' untranslated regions and contained the following sequences:

5'-GACCTCGAGGTTGCTTAAGACTGAAGC-3' (SEQ.ID.NO.: 123)

5'-ATTTCTAGACATATGTAGCTTGTACCG-3' (SEQ.ID.NO.: 124)

Nucleic acid (SEQ.ID.NO.: 125) and amino acid (SEQ.ID.NO.: 126) sequences for human 5HT_{2C} were thereafter determined and verified.

39. V28 (GenBank Accession Number: U20350)

The cDNA for human V28 was generated and cloned into pCMV expression vector as follows: PCR was performed using brain cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1min; and 72°C for 1 min and 20 sec. The 5' PCR primer contained a HindIII site with the sequence:

5'-GGTAAGCTTGGCAGTCCACGCCAGGCCTTC-3' (SEQ.ID.NO.: 127)

and the 3' primer contained an EcoRI site with the sequence:

5'-TCCGAATTCTCTGTAGACACAAGGCTTTGG-3' (SEQ.ID.NO.: 128)

The 1.1 kb PCR fragment was digested with HindIII and EcoRI and cloned into HindIII-EcoRI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 129) and amino acid (SEQ.ID.NO.:

5 130) sequences for human V28 were thereafter determined and verified.

Example 2

PREPARATION OF NON-ENDOGENOUS HUMAN GPCRS

1. Site-Directed Mutagenesis

Mutagenesis based upon the Human GPCR Proline Marker approach disclosed herein was
10 performed on the foregoing endogenous human GPCRs using Transformer Site-Directed
Mutagenesis Kit (Clontech) according to the manufacturer instructions. For this mutagenesis
approach, a Mutation Probe and a Selection Marker Probe (unless otherwise indicated, the probe
of SEQ.ID.NO.: 132 was the same throughout) were utilized, and the sequences of these for the
specified sequences are listed below in Table B (the parenthetical number is the SEQ. ID.NO.).
15 For convenience, the codon mutation incorporated into the human GPCR is also noted, in standard
form:

Table B

Receptor Identifier (Codon Mutation)	Mutation Probe Sequence (5'-3') (SEQ.ID.NO.)	Selection Marker Probe Sequence (5'-3') (SEQ.ID.NO.)
20 GPR1 (F245K)	GATCTCCAGTAGGCATAAGT GGACAATTCTGG (131)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAG (132)
GPR4 (K223A)	AGAAGGCCAAGATCGCGCGG CTGGCCCTCA (133)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
25 GPR5 (V224K)	CGGCGCCACCGCACGAAAA GTCATCTTC	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT

- 51 -

	(134)	
GPR7 (T250K)	GCCAAGAAGCGG <u>GTGAAGTT</u> CCTGGTGGTGGCA (135)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
GPR8 (T259K)	CAGGCGGAAGGTG <u>AAAGTCC</u> TGGTCCTCGT (136)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
30 GPR9 (M254K)	CGGCGCCTGCGGGCC <u>AAGCG</u> GCTGGTGGTGGTG (137)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
GPR9-6 (L241K)	CCAAGCACAAAGCC <u>AAAGAAA</u> GTGACCATCAC (138)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
35 GPR10 (F276K)	GCGCCGGCGCACCC <u>AAATGCT</u> TGCTGGTGGT (139)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
GPR15 (I240K)	CAAAAAGCTGAAGAAATCT <u>A</u> AGAAGATCATCTTTATTGTCG (140)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
GPR17 (V234K)	CAAGACCAAGGCA <u>AAACGCA</u> TGATCGCCAT (141)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
40 GPR18 (I231K)	GTCAAGGAGAAGTCC <u>AAAG</u> GATCATCATC (142)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
GPR20 (M240K)	CGCCGCGTGCGGGCC <u>AAGCA</u> GCTCCTGCTC (143)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
45 GPR21 (A251K)	CCTGATAAGCGCTATA <u>AAAT</u> GGTCCTGTTTCGA (144)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
GPR22 (F312K)	GAAAGACAAAAGAGAGTCA <u>AGAGGATGTCTTTATTG</u> (145)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
GPR24 (T304K)	CGGAGAAAAGAGGGTG <u>AAAC</u> GCACAGCCATCGCC (146)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
50 GPR30 (L258K)	alternate approach; <i>see below</i>	alternate approach; <i>see below</i>
GPR31 (Q221K)	AAGCTTCAGCGGGCC <u>AAGGC</u> ACTGGTCACC (147)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
55 GPR32 (K255A)	CATGCCAACC GGCCCGC <u>GAG</u> GCTGCTGCTGGT (279)	ACCAGCAGCAGCCTCGCGGG CCGGTTGGCATG (280)
GPR40 (A223K)	CGGAAGCTGCGGGCC <u>AAATG</u> GGTGGCCGGC (265)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
GPR41	CAGAGGAGGGTG <u>AAGGGGCT</u> GTTGGCG	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT

	(A223K)	(266)	
	GPR43 (V221K)	GGCGGCGCCGAGCC <u>AAGGGG</u> CTGGCTGTGG (267)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
5	APJ (L247K)	alternate approach; <i>see below</i>	alternate approach; <i>see below</i>
	BLR1 (V258K)	CAGCGGCAGAAAGGCA <u>AAAA</u> GGGTGGCCATC (148)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
	CEPR (L258K)	CGGCAGAAAGGCGA <u>AAGCGCAT</u> GATCCTCGCG (149)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
10	EBI1 (I262K)	GAGCGCAACAAGGCC <u>AAAA</u> AGGTGATCATC (150)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
	EBI2 (L243K)	GGTGTAACAACAAAGGCT <u>AA</u> <u>AAACACAATTATTCTTATT</u> (151)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
15	ETBR-LP2 (N358K)	GAGAGCCAGCTCA <u>AAGAGCAC</u> CGTGGTG (152)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
	GHSR (V262K)	CCACAAGCAAACCA <u>AAGAAAA</u> TGCTGGCTGT (153)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
	GPCR-CNS (N491K)	CTAGAGAGTCAGATGA <u>AGTG</u> TACAGTAGTGGCAC (155)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
20	GPR-NGA (I275K)	CGGACAAAAGTGAAAAC <u>TAA</u> <u>AAAGATGTTCTCATT</u> (156)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
	H9a and H9b (F236K)	GCTGAGGTTGCAAT <u>AAACT</u> AACCATGTTTGTC (157)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
25	HB954 (H265K)	GGGAGGCCGAGCTGA <u>AAAGCC</u> ACCCTGCTC (158)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
	HG38 (V765K)	GGGACTGCTCTATG <u>AAAAAA</u> CACATTGCCCTG (268)	CATCAAGTGATCATGTGCC AAGTACGCCC (154)
	HM74 (I230K)	CAAGATCAAGAGAGCC <u>AAAA</u> CCTTCATCATG (159)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
30	MIG (T273K)	CCGGAGACAAGTGA <u>AAGAAG</u> ATGCTGTTTGTC (160)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
	OGR1 (Q227K)	GCAAGGACCAGATCA <u>AAGCGG</u> CTGGTGCTCA (161)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
35	Serotonin 5HT _{2A} (C322K)	alternate approach; <i>see below</i>	alternate approach; <i>see below</i>
	Serotonin 5HT _{2C} (S310K)	alternate approach; <i>see below</i>	alternate approach; <i>see below</i>

- 53 -

V28 (I230K)	CAAGAAAGCCAAAGCCAAG AAACTGATCCTTCTG (162)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
----------------	---	--------------------------------------

The non-endogenous human GPCRs were then sequenced and the derived and verified nucleic acid and amino acid sequences are listed in the accompanying "Sequence Listing" appendix

5 to this patent document, as summarized in Table C below:

Table C

	Mutated GPCR	Nucleic Acid Sequence Listing	Amino Acid Sequence Listing
	GPR1 (F245K)	SEQ.ID.NO.: 163	SEQ.ID.NO.: 164
10	GPR4 (K223A)	SEQ.ID.NO.: 165	SEQ.ID.NO.: 166
	GPR5 (V224K)	SEQ.ID.NO.: 167	SEQ.ID.NO.: 168
15	GPR7 (T250K)	SEQ.ID.NO.: 169	SEQ.ID.NO.: 170
	GPR8 (T259K)	SEQ.ID.NO.: 171	SEQ.ID.NO.: 172
	GPR9 (M254K)	SEQ.ID.NO.: 173	SEQ.ID.NO.: 174
20	GPR9-6 (L241K)	SEQ.ID.NO.: 175	SEQ.ID.NO.: 176
	GPR10 (F276K)	SEQ.ID.NO.: 177	SEQ.ID.NO.: 178
25	GPR15 (I240K)	SEQ.ID.NO.: 179	SEQ.ID.NO.: 180
	GPR17 (V234K)	SEQ.ID.NO.: 181	SEQ.ID.NO.: 182
	GPR18 (I231K)	SEQ.ID.NO.: 183	SEQ.ID.NO.: 184
30	GPR20 (M240K)	SEQ.ID.NO.: 185	SEQ.ID.NO.: 186
	GPR21 (A251K)	SEQ.ID.NO.: 187	SEQ.ID.NO.: 188
35	GPR22 (F312K)	SEQ.ID.NO.: 189	SEQ.ID.NO.: 190
	GPR24 (T304K))	SEQ.ID.NO.: 191	SEQ.ID.NO.: 192
	GPR30	SEQ.ID.NO.: 193	SEQ.ID.NO.: 194

	(L258K)		
	GPR31 (Q221K)	SEQ.ID.NO.: 195	SEQ.ID.NO.: 196
5	GPR32 (K255A)	SEQ.ID.NO.: 269	SEQ.ID.NO.: 270
	GPR40 (A223K)	SEQ.ID.NO.: 271	SEQ.ID.NO.: 272
	GPR41 (A223K)	SEQ.ID.NO.: 273	SEQ.ID.NO.: 274
10	GPR43 (V221K)	SEQ.ID.NO.: 275	SEQ.ID.NO.: 276
	APJ (L247K)	SEQ.ID.NO.: 197	SEQ.ID.NO.: 198
15	BLR1 (V258K)	SEQ.ID.NO.: 199	SEQ.ID.NO.: 200
	CEPR (L258K)	SEQ.ID.NO.: 201	SEQ.ID.NO.: 202
	EBI1 (I262K)	SEQ.ID.NO.: 203	SEQ.ID.NO.: 204
20	EBI2 (L243K)	SEQ.ID.NO.: 205	SEQ.ID.NO.: 206
	ETBR-LP2 (N358K)	SEQ.ID.NO.: 207	SEQ.ID.NO.: 208
25	GHSR (V262K)	SEQ.ID.NO.: 209	SEQ.ID.NO.: 210
	GPCR-CNS (N491K)	SEQ.ID.NO.: 211	SEQ.ID.NO.: 212
	GPR-NGA (I275K)	SEQ.ID.NO.: 213	SEQ.ID.NO.: 214
30	H9a (F236K)	SEQ.ID.NO.: 215	SEQ.ID.NO.: 216
	H9b (F236K)	SEQ.ID.NO.: 217	SEQ.ID.NO.: 218
35	HB954 (H265K)	SEQ.ID.NO.: 219	SEQ.ID.NO.: 220
	HG38 (V765K)	SEQ.ID.NO.: 277	SEQ.ID.NO.: 278
	HM74 (I230K)	SEQ.ID.NO.: 221	SEQ.ID.NO.: 222
40	MIG (T273K)	SEQ.ID.NO.: 223	SEQ.ID.NO.: 224
	OGR1 (Q227K)	SEQ.ID.NO.: 225	SEQ.ID.NO.: 226
45	Serotonin 5HT _{2A} (C322K)	SEQ.ID.NO.: 227	SEQ.ID.NO.: 228
	Serotonin 5HT _{2C} (S310K)	SEQ.ID.NO.: 229	SEQ.ID.NO.: 230
	V28 (I230K)	SEQ.ID.NO.: 231	SEQ.ID.NO.: 232

- 55 -

2. Alternate Mutation Approaches for Employment of the Proline Marker Algorithm: APJ; Serotonin 5HT_{2A}; Serotonin 5HT_{2C}; and GPR30

Although the above site-directed mutagenesis approach is particularly preferred, other approaches can be utilized to create such mutations; those skilled in the art are readily credited with selecting approaches to mutating a GPCR that fits within the particular needs of the artisan.

a. APJ

Preparation of the non-endogenous, human APJ receptor was accomplished by mutating L247K. Two oligonucleotides containing this mutation were synthesized:

5'-GGCTTAAGAGCATCATCGTGGTGCTGGTG-3' (SEQ.ID.NO.: 233)

10 5'-GTCACCACCAGCACCACGATGATGCTCTTAAGCC-3' (SEQ.ID.NO.: 234)

The two oligonucleotides were annealed and used to replace the NaeI-BstEII fragment of human, endogenous APJ to generate the non-endogenous, version of human APJ.

b. Serotonin 5HT_{2A}

cDNA containing the point mutation C322K was constructed by utilizing the restriction enzyme site Sph I which encompasses amino acid 322. A primer containing the C322K mutation:

5'-CAAAGAAAGTACTGGGCATCGTCTTCTTCCT-3' (SEQ.ID.NO: 235)

was used along with the primer from the 3' untranslated region of the receptor:

5'-TGCTCTAGATTCCAGATAGGTGAAAA CTTG-3' (SEQ.ID.NO.: 236)

to perform PCR (under the conditions described above). The resulting PCR fragment was then used to replace the 3' end of endogenous 5HT_{2A} cDNA through the T4 polymerase blunted Sph I site.

c. Serotonin 5HT_{2C}

The cDNA containing a S310K mutation was constructed by replacing the Sty I restriction fragment containing amino acid 310 with synthetic double stranded oligonucleotides that encode

the desired mutation. The sense strand sequence utilized had the following sequence:

5'-CTAGGGGCACCATGCAGGCTATCAACAATGAAAGAAAAGCTAAGAAAGTC-3'

(SEQ. ID.NO.: 237)

and the antisense strand sequence utilized had the following sequence:

5'-CAAGGACTTTCTTAGCTTTTCTTTCATTGTTGATAGCCTGCATGGTGCCC-3' (SEQ.

ID. NO.: 238)

d. GPR30

Prior to generating non-endogenous GPR30, several independent pCR2.1/GPR30 isolates were sequenced in their entirety in order to identify clones with no PCR-generated mutations. A
10 clone having no mutations was digested with EcoRI and the endogenous GPR30 cDNA fragment was transferred into the CMV-driven expression plasmid pCI-neo (Promega), by digesting pCI-Neo with EcoRI and subcloning the EcoRI-liberated GPR30 fragment from pCR2.1/GPR30, to generate pCI/GPR30. Thereafter, the leucine at codon 258 was mutated to a lysine using a Quick-Change™ Site-Directed Mutagenesis Kit (Stratagene, #200518), according to manufacturer's
15 instructions, and the following primers:

5'-CGGCGGCAGAAGGCGAAACGCATGATCCTCGCGGT-3' (SEQ.ID.NO.: 239) and

5'-ACCGCGAGGATCATGCGTTTCGCCTTCTGC CGCCG-3' (SEQ.ID.NO.: 240)

Example 3

Receptor (Endogenous and Mutated) Expression

20

Although a variety of cells are available to the art for the expression of proteins, it is most preferred that mammalian cells be utilized. The primary reason for this is predicated upon practicalities, *i.e.*, utilization of, *e.g.*, yeast cells for the expression of a GPCR, while possible,

- 57 -

introduces into the protocol a non-mammalian cell which may not (indeed, in the case of yeast, does not) include the receptor-coupling, genetic-mechanism and secretory pathways that have evolved for mammalian systems – thus, results obtained in non-mammalian cells, while of potential use, are not as preferred as that obtained from mammalian cells. Of the mammalian cells, COS-7, 293 and 293T cells are particularly preferred, although the specific mammalian cell utilized can be predicated upon the particular needs of the artisan.

Unless otherwise noted herein, the following protocol was utilized for the expression of the endogenous and non-endogenous human GPCRs. Table D lists the mammalian cell and number utilized (per 150mm plate) for GPCR expression.

Table D

Receptor Name (Endogenous or Non-Endogenous)	Mammalian Cell (Number Utilized)
GPR17	293 (2×10^4)
GPR30	293 (4×10^4)
APJ	COS-7 (5×10^6)
ETBR-LP2	293 (1×10^7)
	293T (1×10^7)
GHSR	293 (1×10^7)
	293T (1×10^7)
MIG	293 (1×10^7)
Serotonin 5HT _{2A}	293T (1×10^7)
Serotonin 5HT _{2c}	293T (1×10^7)

On day one, mammalian cells were plated out. On day two, two reaction tubes were prepared (the proportions to follow for each tube are per plate): tube A was prepared by mixing 20µg DNA (e.g., pCMV vector; pCMV vector with endogenous receptor cDNA, and pCMV vector with non-endogenous receptor cDNA.) in 1.2ml serum free DMEM (Irvine Scientific,

Irvine, CA); tube B was prepared by mixing 120µl lipofectamine (Gibco BRL) in 1.2ml serum free DMEM. Tubes A and B were then admixed by inversions (several times), followed by incubation at room temperature for 30-45min. The admixture is referred to as the "transfection mixture". Plated cells were washed with 1XPBS, followed by addition of 10ml serum free
5 DMEM. 2.4ml of the transfection mixture was then added to the cells, followed by incubation for 4hrs at 37°C/5% CO₂. The transfection mixture was then removed by aspiration, followed by the addition of 25ml of DMEM/10% Fetal Bovine Serum. Cells were then incubated at 37°C/5% CO₂. After 72hr incubation, cells were then harvested and utilized for analysis.

1. **Gi-Coupled Receptors: Co-Transfection with Gs-Coupled Receptors**

10 In the case of GPR30, it has been determined that this receptor couples the G protein Gi. Gi is known to inhibit the enzyme adenylyl cyclase, which is necessary for catalyzing the conversion of ATP to cAMP. Thus, a non-endogenous, constitutively activated form of GPR30 would be expected to be associated with decreased levels of cAMP. Assay confirmation of a non-endogenous, constitutively activated form of GPR30 directly via measurement of decreasing levels
15 of cAMP, while viable, can be preferably measured by cooperative use of a Gs-coupled receptor. For example, a receptor that is Gs-coupled will stimulate adenylyl cyclase, and thus will be associated with an increase in cAMP. The assignee of the present application has discovered that the orphan receptor GPR6 is an endogenous, constitutively activated GPCR. GPR6 couples to the Gs protein. Thus when co-transfected, one can readily verify that a putative GPR30-mutation
20 leads to constitutive activation thereof: *i.e.*, an endogenous, constitutively activated GPR6/endogenous, non-constitutively activated GPR30 cell will evidence an elevated level of cAMP when compared with an endogenous, constitutively active GPR6/non-endogenous, constitutively activated GPR30 (the latter evidencing a comparatively lower level of cAMP).

- 59 -

Assays that detect cAMP can be utilized to determine if a candidate compound is *e.g.*, an inverse agonist to a Gs-associated receptor (*i.e.*, such a compound would decrease the levels of cAMP) or a Gi-associated receptor (or a Go-associated receptor) (*i.e.*, such a candidate compound would increase the levels of cAMP). A variety of approaches known in the art for measuring cAMP can be utilized; a preferred approach relies upon the use of anti-cAMP antibodies. Another approach, and most preferred, utilizes a whole cell second messenger reporter system assay. Promoters on genes drive the expression of the proteins that a particular gene encodes. Cyclic AMP drives gene expression by promoting the binding of a cAMP-responsive DNA binding protein or transcription factor (CREB) which then binds to the promoter at specific sites called cAMP response elements and drives the expression of the gene. Reporter systems can be constructed which have a promoter containing multiple cAMP response elements before the reporter gene, *e.g.*, β -galactosidase or luciferase. Thus, an activated receptor such as GPR6 causes the accumulation of cAMP which then activates the gene and expression of the reporter protein. Most preferably, 293 cells are co-transfected with GPR6 (or another Gs-linked receptor) and GPR30 (or another Gi-linked receptor) plasmids, preferably in a 1:1 ratio, most preferably in a 1:4 ratio. Because GPR6 is an endogenous, constitutively active receptor that stimulates the production of cAMP, GPR6 strongly activates the reporter gene and its expression. The reporter protein such as β -galactosidase or luciferase can then be detected using standard biochemical assays (Chen et al. 1995). Co-transfection of endogenous, constitutively active GPR6 with endogenous, non-constitutively active GPR30 evidences an increase in the luciferase reporter protein. Conversely, co-transfection of endogenous, constitutively active GPR6 with non-endogenous, constitutively active GPR30 evidences a drastic decrease in expression of luciferase. Several reporter plasmids are known and available in the art for measuring a second messenger assay. It is considered well within the

skilled artisan to determine an appropriate reporter plasmid for a particular gene expression based primarily upon the particular need of the artisan. Although a variety of cells are available for expression, mammalian cells are most preferred, and of these types, 293 cells are most preferred. 293 cells were transfected with the reporter plasmid pCRE-Luc/GPR6 and non-endogenous, 5 constitutively activated GPR30 using a Mammalian Transfection™ Kit (Stratagene, #200285) CaPO₄ precipitation protocol according to the manufacturer's instructions (*see*, 28 Genomics 347 (1995) for the published endogenous GPR6 sequence). The precipitate contained 400ng reporter, 80ng CMV-expression plasmid (having a 1:4 GPR6 to endogenous GPR30 or non-endogenous GPR30 ratio) and 20ng CMV-SEAP (a transfection control plasmid encoding secreted alkaline 10 phosphatase). 50% of the precipitate was split into 3 wells of a 96-well tissue culture dish (containing 4×10^4 cells/well); the remaining 50% was discarded. The following morning, the media was changed. 48 hr after the start of the transfection, cells were lysed and examined for luciferase activity using a Lucite™ Kit (Packard, Cat. # 6016911) and Trilux 1450 Microbeta™ liquid scintillation and luminescence counter (Wallac) as per the vendor's instructions. The data 15 were analyzed using GraphPad Prism 2.0a (GraphPad Software Inc.).

With respect to GPR17, which has also been determined to be Gi-linked, a modification of the foregoing approach was utilized, based upon, *inter alia*, use of another Gs-linked endogenous receptor, GPR3 (*see* 23 Genomics 609 (1994) and 24 Genomics 391 (1994)). Most preferably, 293 cells are utilized. These cells were plated-out on 96 well plates at a density of 20 $\times 10^4$ cells per well and were transfected using Lipofectamine Reagent (BRL) the following day according to manufacturer instructions. A DNA/lipid mixture was prepared for each 6-well transfection as follows: 260ng of plasmid DNA in 100μl of DMEM were gently mixed with 2μl of lipid in 100μl of DMEM (the 260ng of plasmid DNA consisted of 200ng of a 8xCRE-Luc

- 61 -

reporter plasmid (*see below*), 50ng of pCMV comprising endogenous receptor or non-endogenous receptor or pCMV alone, and 10ng of a GPRS expression plasmid (GPRS in pcDNA3 (Invitrogen)). The 8XCRE-Luc reporter plasmid was prepared as follows: vector SRIF- β -gal was obtained by cloning the rat somatostatin promoter (-71/+51) at BglV-HindIII site in the p β gal-Basic Vector (Clontech). Eight (8) copies of cAMP response element were obtained by PCR from an adenovirus template AdpCF126CCRE8 (*see 7 Human Gene Therapy 1883 (1996)*) and cloned into the SRIF- β -gal vector at the Kpn-BglV site, resulting in the 8xCRE- β -gal reporter vector. The 8xCRE-Luc reporter plasmid was generated by replacing the beta-galactosidase gene in the 8xCRE- β -gal reporter vector with the luciferase gene obtained from the pGL3-basic vector (Promega) at the HindIII-BamHI site. Following 30min. incubation at room temperature, the DNA/lipid mixture was diluted with 400 μ l of DMEM and 100 μ l of the diluted mixture was added to each well. 100 μ l of DMEM with 10% FCS were added to each well after a 4hr incubation in a cell culture incubator. The next morning the transfected cells were changed with 200 μ l/well of DMEM with 10% FCS. Eight (8) hours later, the wells were changed to 100 μ l /well of DMEM without phenol red, after one wash with PBS. Luciferase activity were measured the next day using the LucLite™ reporter gene assay kit (Packard) following manufacturer instructions and read on a 1450 MicroBeta™ scintillation and luminescence counter (Wallac).

Figure 4 evidences that constitutively active GPR30 inhibits GPR6-mediated activation of CRE-Luc reporter in 293 cells. Luciferase was measured at about 4.1 relative light units in the expression vector pCMV. Endogenous GPR30 expressed luciferase at about 8.5 relative light units, whereas the non-endogenous, constitutively active GPR30 (L258K), expressed luciferase at about 3.8 and 3.1 relative light units, respectively. Co-transfection of endogenous GPR6 with endogenous GPR30, at a 1:4 ratio, drastically increased luciferase

expression to about 104.1 relative light units. Co-transfection of endogenous GPR6 with non-endogenous GPR30 (L258K), at the same ratio, drastically decreased the expression, which is evident at about 18.2 and 29.5 relative light units, respectively. Similar results were observed with respect to GPR17 with respect to co-transfection with GPR3, as set forth in Figure 5.

Example 3

ASSAYS FOR DETERMINATION OF CONSTITUTIVE ACTIVITY OF NON-ENDOGENOUS GPCRS

A. Membrane Binding Assays

1. [³⁵S]GTPγS Assay

When a G protein-coupled receptor is in its active state, either as a result of ligand binding or constitutive activation, the receptor couples to a G protein and stimulates the release of GDP and subsequent binding of GTP to the G protein. The alpha subunit of the G protein-receptor complex acts as a GTPase and slowly hydrolyzes the GTP to GDP, at which point the receptor normally is deactivated. Constitutively activated receptors continue to exchange GDP for GTP. The non-hydrolyzable GTP analog, [³⁵S]GTPγS, can be utilized to demonstrate enhanced binding of [³⁵S]GTPγS to membranes expressing constitutively activated receptors. The advantage of using [³⁵S]GTPγS binding to measure constitutive activation is that: (a) it is generically applicable to all G protein-coupled receptors; (b) it is proximal at the membrane surface making it less likely to pick-up molecules which affect the intracellular cascade.

The assay utilizes the ability of G protein coupled receptors to stimulate [³⁵S]GTPγS binding to membranes expressing the relevant receptors. The assay can, therefore, be used in the direct identification method to screen candidate compounds to known, orphan and constitutively activated G protein-coupled receptors. The assay is generic and has application

- 63 -

to drug discovery at all G protein-coupled receptors.

The [^{35}S]GTP γ S assay was incubated in 20 mM HEPES and between 1 and about 20mM MgCl_2 (this amount can be adjusted for optimization of results, although 20mM is preferred) pH 7.4, binding buffer with between about 0.3 and about 1.2 nM [^{35}S]GTP γ S (this amount can be adjusted for optimization of results, although 1.2 is preferred) and 12.5 to 75 μg membrane protein (e.g. COS-7 cells expressing the receptor; this amount can be adjusted for optimization, although 75 μg is preferred) and 1 μM GDP (this amount can be changed for optimization) for 1 hour. Wheatgerm agglutinin beads (25 μl ; Amersham) were then added and the mixture was incubated for another 30 minutes at room temperature. The tubes were then centrifuged at 1500 x g for 5 minutes at room temperature and then counted in a scintillation counter.

A less costly but equally applicable alternative has been identified which also meets the needs of large scale screening. Flash platesTM and WallacTM scintistrips may be utilized to format a high throughput [^{35}S]GTP γ S binding assay. Furthermore, using this technique, the assay can be utilized for known GPCRs to simultaneously monitor tritiated ligand binding to the receptor at the same time as monitoring the efficacy via [^{35}S]GTP γ S binding. This is possible because the Wallac beta counter can switch energy windows to look at both tritium and ^{35}S -labeled probes. This assay may also be used to detect other types of membrane activation events resulting in receptor activation. For example, the assay may be used to monitor ^{32}P phosphorylation of a variety of receptors (both G protein coupled and tyrosine kinase receptors). When the membranes are centrifuged to the bottom of the well, the bound [^{35}S]GTP γ S or the ^{32}P -phosphorylated receptor will activate the scintillant which is coated of the wells. Scinti[®] strips (Wallac) have been used to demonstrate this principle. In addition, the assay also has utility for measuring ligand binding to receptors using radioactively labeled ligands. In a similar manner, when the radiolabeled bound

ligand is centrifuged to the bottom of the well, the scintistrip label comes into proximity with the radiolabeled ligand resulting in activation and detection.

Representative results of graph comparing Control (pCMV), Endogenous APJ and Non-Endogenous APJ, based upon the foregoing protocol, are set forth in Figure 6.

5 2. Adenylyl Cyclase

A Flash Plate™ Adenylyl Cyclase kit (New England Nuclear; Cat. No. SMP004A) designed for cell-based assays was modified for use with crude plasma membranes. The Flash Plate wells contain a scintillant coating which also contains a specific antibody recognizing cAMP. The cAMP generated in the wells was quantitated by a direct competition for binding of
10 radioactive cAMP tracer to the cAMP antibody. The following serves as a brief protocol for the measurement of changes in cAMP levels in membranes that express the receptors.

Transfected cells were harvested approximately three days after transfection. Membranes were prepared by homogenization of suspended cells in buffer containing 20mM HEPES, pH 7.4 and 10mM MgCl₂. Homogenization was performed on ice using a Brinkman Polytron™ for
15 approximately 10 seconds. The resulting homogenate was centrifuged at 49,000 X g for 15 minutes at 4°C. The resulting pellet was then resuspended in buffer containing 20mM HEPES, pH 7.4 and 0.1 mM EDTA, homogenized for 10 seconds, followed by centrifugation at 49,000 X g for 15 minutes at 4°C. The resulting pellet can be stored at -80°C until utilized. On the day of measurement, the membrane pellet was slowly thawed at room temperature, resuspended in buffer
20 containing 20mM HEPES, pH 7.4 and 10mM MgCl₂ (these amounts can be optimized, although the values listed herein are preferred), to yield a final protein concentration of 0.60mg/ml (the resuspended membranes were placed on ice until use).

cAMP standards and Detection Buffer (comprising 2 µCi of tracer [¹²⁵I cAMP (100 µl] to

- 65 -

11 ml Detection Buffer) were prepared and maintained in accordance with the manufacturer's instructions. Assay Buffer was prepared fresh for screening and contained 20mM HEPES, pH 7.4, 10mM $MgCl_2$, 20mM (Sigma), 0.1 units/ml creatine phosphokinase (Sigma), 50 μ M GTP (Sigma), and 0.2 mM ATP (Sigma); Assay Buffer can be stored on ice until utilized. The assay was initiated by addition of 50ul of assay buffer followed by addition of 50ul of membrane suspension to the NEN Flash Plate. The resultant assay mixture is incubated for 60 minutes at room temperature followed by addition of 100ul of detection buffer. Plates are then incubated an additional 2-4 hours followed by counting in a Wallac MicroBeta scintillation counter. Values of cAMP/well are extrapolated from a standard cAMP curve which is contained within each assay plate. The foregoing assay was utilized with respect to analysis of MIG.

B. Reporter-Based Assays

1. CREB Reporter Assay (Gs-associated receptors)

A method to detect Gs stimulation depends on the known property of the transcription factor CREB, which is activated in a cAMP-dependent manner. A PathDetect CREB Reporting System (Stratagene, Catalogue # 219010) was utilized to assay for Gs coupled activity in 293 or 293T cells. Cells were transfected with the plasmids components of this above system and the indicated expression plasmid encoding endogenous or mutant receptor using a Mammalian Transfection Kit (Stratagene, Catalogue #200285) according to the manufacturer's instructions. Briefly, 400 ng pFR-Luc (luciferase reporter plasmid containing Gal4 recognition sequences), 40 ng pFA2-CREB (Gal4-CREB fusion protein containing the Gal4 DNA-binding domain), 80 ng CMV-receptor expression plasmid (comprising the receptor) and 20 ng CMV-SEAP (secreted alkaline phosphatase expression plasmid; alkaline phosphatase activity is measured in the media of transfected cells to control for variations in

transfection efficiency between samples) were combined in a calcium phosphate precipitate as per the Kit's instructions. Half of the precipitate was equally distributed over 3 wells in a 96-well plate, kept on the cells overnight, and replaced with fresh medium the following morning. Forty-eight (48) hr after the start of the transfection, cells were treated and assayed for luciferase activity as set forth with respect to the GPR30 system, above. This assay was used with respect to GHSR.

2. AP1 reporter assay (Gq-associated receptors)

A method to detect Gq stimulation depends on the known property of Gq-dependent phospholipase C to cause the activation of genes containing AP1 elements in their promoter. A Pathdetect AP-1 cis-Reporting System (Stratagene, Catalogue # 219073) was utilized following the protocol set forth above with respect to the CREB reporter assay, except that the components of the calcium phosphate precipitate were 410 ng pAP1-Luc, 80 ng receptor expression plasmid, and 20 ng CMV-SEAP. This assay was used with respect to ETBR-LP2

C. Intracellular IP3 Accumulation Assay

On day 1, cells comprising the serotonin receptors (endogenous and mutated) were plated onto 24 well plates, usually 1×10^5 cells/well. On day 2 cells were transfected by firstly mixing 0.25ug DNA in 50 ul serumfree DMEM/well and 2 ul lipofectamine in 50 ul serumfree DMEM/well. The solutions were gently mixed and incubated for 15-30 min at room temperature. Cells were washed with 0.5 ml PBS and 400 ul of serum free media was mixed with the transfection media and added to the cells. The cells were then incubated for 3-4 hrs at $37^\circ\text{C}/5\%\text{CO}_2$ and then the transfection media was removed and replaced with 1ml/well of regular growth media. On day 3 the cells were labeled with ^3H -myo-inositol. Briefly, the media was removed the cells were washed with 0.5 ml PBS. Then 0.5 ml inositol-free/serumfree media (GIBCO BRL) was added/well with 0.25 μCi of ^3H -myo-inositol / well

- 67 -

and the cells were incubated for 16-18 hrs o/n at 37°C/5%CO₂. On Day 4 the cells were washed with 0.5 ml PBS and 0.45 ml of assay medium was added containing inositol-free/serum free media 10 µM pargyline 10 mM lithium chloride or 0.4 ml of assay medium and 50 µl of 10x ketanserin (ket) to final concentration of 10µM. The cells were then
5 incubated for 30 min at 37°C. The cells were then washed with 0.5 ml PBS and 200 µl of fresh/ice cold stop solution (1M KOH; 18 mM Na-borate; 3.8 mM EDTA) was added/well. The solution was kept on ice for 5-10 min or until cells were lysed and then neutralized by 200 µl of fresh/ice cold neutralization sol. (7.5 % HCL). The lysate was then transferred into 1.5 ml eppendorf tubes and 1 ml of chloroform/methanol (1:2) was added/tube. The solution
10 was vortexed for 15 sec and the upper phase was applied to a Biorad AG1-X8 anion exchange resin (100-200 mesh). Firstly, the resin was washed with water at 1:1.25 W/V and 0.9 ml of upper phase was loaded onto the column. The column was washed with 10 mls of 5 mM myo-inositol and 10 ml of 5 mM Na-borate/60mM Na-formate. The inositol tris phosphates were eluted into scintillation vials containing 10 ml of scintillation cocktail with
15 2 ml of 0.1 M formic acid/ 1 M ammonium formate. The columns were regenerated by washing with 10 ml of 0.1 M formic acid/3M ammonium formate and rinsed twice with dd H₂O and stored at 4°C in water.

Figure 7 provides an illustration of IP3 production from the human 5-HT_{2A} receptor that incorporates the C322K mutation. While these results evidence that the Proline Mutation
20 Algorithm approach constitutively activates this receptor, for purposes of using such a receptor for screening for identification of potential therapeutics, a more robust difference would be preferred. However, because the activated receptor can be utilized for understanding and elucidating the role of constitutive activation and for the identification of compounds that

can be further examined, we believe that this difference is itself useful in differentiating between the endogenous and non-endogenous versions of the human 5HT_{2A} receptor.

D. Result Summary

The results for the GPCRs tested are set forth in Table E where the Per-Cent Increase indicates the percentage difference in results observed for the non-endogenous GPCR as compared to the endogenous GPCR; these values are followed by parenthetical indications as to the type of assay utilized. Additionally, the assay system utilized is parenthetically listed (and, in cases where different Host Cells were used, both are listed). As these results indicate, a variety of assays can be utilized to determine constitutive activity of the non-endogenous versions of the human GPCRs. Those skilled in the art, based upon the foregoing and with reference to information available to the art, are credited with the ability to select and/or maximize a particular assay approach that suits the particular needs of the investigator.

Table E

Receptor Identifier (Codon Mutation)	Per-Cent Difference
GPR17 (V234K)	74.5 (CRE-Luc)
GPR30 (L258K)	71.6 (CREB)
APJ (L247K)	49.0 (GTP γ S)
ETBR-LP2 (N358K)	48.4(AP1-Luc - 293) 61.1(AP1-Luc - 293T)
GHSR (V262K)	58.9(CREB - 293) 35.6(CREB - 293T)

- 69 -

MIG (I230K)	39 (cAMP)
Serotonin 5HT _{2A} (C322K)	33.2 (IP ₃)
Serotonin 5HT _{2C} (S310K)	39.1 (IP ₃)

Example 6**Tissue Distribution of Endogenous Orphan GPCRs**

Using a commercially available human-tissue dot-blot format, endogenous orphan GPCRs were probed for a determination of the areas where such receptors are localized. Except as indicate below, the entire receptor cDNA (radiolabelled) was used as the probe: radiolabeled probe was generated using the complete receptor cDNA (excised from the vector) using a Prime-It II™ Random Primer Labeling Kit (Stratagene, #300385), according to manufacturer's instructions. A human RNA Master Blot™ (Clontech, #7770-1) was hybridized with the GPCR radiolabeled probe and washed under stringent conditions according manufacturer's instructions. The blot was exposed to Kodak BioMax Autoradiography film overnight at -80°C.

Representative dot-blot format results are presented in Figure 8 for GPR1 (8A), GPR30 (8B), and APJ (8C), with results being summarized for all receptors in Table F

Table F

GPCR	Tissue Distribution (highest levels, relative to other tissues in the dot-bl t)
GPR1	Placenta, Ovary, Adrenal

	GPR4	Broad; highest in Heart, Lung, Adrenal, Thyroid, Spinal Cord
	GPR5	Placenta, Thymus, Fetal Thymus Lesser levels in spleen, fetal spleen
	GPR7	Liver, Spleen, Spinal Cord, Placenta
	GPR8	No expression detected
5	GPR9-6	Thymus, Fetal Thymus Lesser levels in Small Intestine
	GPR18	Spleen, Lymph Node, Fetal Spleen, Testis
	GPR20	Broad
	GPR21	Broad; very low abundance
	GPR22	Heart, Fetal Heart Lesser levels in Brain
10	GPR30	Stomach
	GPR31	Broad
	BLR1	Spleen
	CEPR	Stomach, Liver, Thyroid, Putamen
	EBI1	Pancreas Lesser levels in Lymphoid Tissues
15	EBI2	Lymphoid Tissues, Aorta, Lung, Spinal Cord
	ETBR-LP2	Broad; Brain Tissue
	GPCR-CNS	Brain Lesser levels in Testis, Placenta
	GPR-NGA	Pituitary Lesser levels in Brain
	H9	Pituitary
20	HB954	Aorta, Cerebellum Lesser levels in most other tissues
	HM74	Spleen, Leukocytes, Bone marrow, Mammary Glands, Lung, Trachea
	MIG	Low levels in Kidney, Liver, Pancreas, Lung, Spleen
	ORG1	Pituitary, Stomach, Placenta
	V28	Brain, Spleen, Peripheral Leukocytes

25 Based upon the foregoing information, it is noted that human GPCRs can also be assessed for distribution in diseased tissue; comparative assessments between "normal" and diseased tissue can then be utilized to determine the potential for over-expression or under-expression of a particular receptor in a diseased state. In those circumstances where it is desirable to utilize the non-endogenous versions of the human GPCRs for the purpose of screening to directly identify

- 71 -

candidate compounds of potential therapeutic relevance, it is noted that inverse agonists are useful in the treatment of diseases and disorders where a particular human GPCR is over-expressed, whereas agonists or partial agonists are useful in the treatment of diseases and disorders where a particular human GPCR is under-expressed.

5 As desired, more detailed, cellular localization of the receptors, using techniques well-known to those in the art (*e.g.*, in-situ hybridization) can be utilized to identify particular cells within these tissues where the receptor of interest is expressed.

It is intended that each of the patents, applications, and printed publications mentioned in this patent document be hereby incorporated by reference in their entirety.

10 As those skilled in the art will appreciate, numerous changes and modifications may be made to the preferred embodiments of the invention without departing from the spirit of the invention. It is intended that all such variations fall within the scope of the invention.

Although a variety of expression vectors are available to those in the art, for purposes of utilization for both the endogenous and non-endogenous human GPCRs, it is most preferred that
15 the vector utilized be pCMV. This vector has been deposited with the American Type Culture Collection (ATCC) on October 13, 1998 (10801 University Blvd., Manassas, VA 20110-2209 USA) under the provisions of the Budapest Treaty for the International Recognition of the Deposit of Microorganisms for the Purpose of patent Procedure. The vector was tested by the ATCC on _____, 1998 and determined to be viable on _____, 1998. The ATCC has assigned
20 the following deposit number to pCMV: _____.

CLAIMS

What is claimed is:

1. A constitutively active, non-endogenous version of an endogenous human orphan G protein-coupled receptor (GPCR) comprising the following amino acid residues (carboxy-terminus to amino-terminus orientation) transversing the transmembrane-6 (TM6) and intracellular loop-3 (IC3) regions of the non-endogenous GPCR:

$P^I AA_{15} X$

wherein:

- (1) P^I is an amino acid residue located within the TM6 region of the non-endogenous GPCR, where P^I is selected from the group consisting of (i) the endogenous orphan GPCR proline residue, and (ii) a non-endogenous amino acid residue other than proline;
- (2) AA_{15} are 15 amino acid residues selected from the group consisting of (a) the 15 endogenous amino acid residues of the endogenous orphan GPCR, (b) 15 non-endogenous amino acid residues, and (c) a combination of 15 amino acid residues, the combination comprising at least one endogenous amino acid residue of the endogenous orphan GPCR and at least one non-endogenous amino acid residue, excepting that none of the 15 endogenous amino acid residues that are positioned within the TM6 region of the GPCR is proline; and
- (2) X is a non-endogenous amino acid residue located within the IC3 region of said non-endogenous GPCR.
2. The non-endogenous human GPCR of claim 1 wherein P^I is the endogenous proline

- 73 -

residue.

3. The non-endogenous human GPCR of claim 1 wherein P¹ is a non-endogenous amino acid residue other than a proline residue.
4. The non-endogenous human GPCR of claim 1 wherein AA₁₅ are the 15 endogenous amino acid residues of the endogenous GPCR.
5. The non-endogenous human GPCR of claim 1 wherein X is selected from the group consisting of lysine, histidine, arginine and alanine residues, excepting that when the endogenous amino acid in position X of said endogenous human GPCR is lysine, X is selected from the group consisting of histidine, arginine and alanine.
- 10 6. The non-endogenous human GPCR of claim 1 wherein X is a lysine residue, excepting that when the endogenous amino acid in position X of said endogenous human GPCR is lysine, X is an amino acid other than lysine.
7. The non-endogenous human GPCR of claim 4 wherein X is a lysine residue, excepting that when the endogenous amino acid in position X of said endogenous human GPCR is lysine, X is an amino acid other than lysine.
- 15 8. The non-endogenous, human GPCR of claim 1 wherein P¹ is a proline residue and X is a lysine residue, excepting that when the endogenous amino acid in position X of said endogenous human GPCR is lysine, X is an amino acid other than lysine.
9. A host cell comprising the non-endogenous human GPCR of claim 1.
- 20 10. The material of claim 9 wherein said host cell is of mammalian origin.
11. The non-endogenous human GPCR of claim 1 in a purified and isolated form.
12. A nucleic acid sequence encoding a constitutively active, non-endogenous version of an endogenous human orphan G protein-coupled receptor (GPCR) comprising the following

nucleic acid sequence region transversing the transmembrane-6 (TM6) and intracellular loop-3 (IC3) regions of the orphan GPCR:



wherein:

- 5 (1) P^{codon} is a nucleic acid encoding region within the TM6 region of the non-endogenous GPCR, where P^{codon} encodes an amino acid selected from the group consisting of (i) the endogenous GPCR proline residue, and (ii) a non-endogenous amino acid residue other than proline;
 - 10 (2) $(\text{AA-codon})_{15}$ are 15 codons encoding 15 amino acid residues selected from the group consisting of (a) the 15 endogenous amino acid residues of the endogenous orphan GPCR, (b) 15 non-endogenous amino acid residues, and (c) a combination of 15 amino acid residues, the combination comprising at least one endogenous amino acid residue of the endogenous orphan GPCR and at least one non-endogenous amino acid residue, excepting that none of the 15 endogenous amino acid residues that are positioned within the TM6 region of the orphan GPCR is proline; and
 - 15 (3) X_{codon} is a nucleic acid encoding region residue located within the IC3 region of said non-endogenous human GPCR, where X_{codon} encodes a non-endogenous amino acid.
 - 20
13. The nucleic acid sequence of claim 12 wherein P^{codon} encodes an endogenous proline residue.
 14. The nucleic acid sequence of claim 12 wherein P^{codon} encodes a non-endogenous

- 75 -

- amino acid residue other than a proline residue.
15. The nucleic acid sequence of claim 12 wherein X_{codon} encodes a non-endogenous amino acid selected from the group consisting of lysine, histidine, arginine and alanine, excepting that when the endogenous amino acid in position X of said endogenous human GPCR is lysine, X_{codon} encodes an amino acid selected from the group consisting of histidine, arginine and alanine.
16. The nucleic acid sequence of claim 13 wherein X_{codon} encodes a non-endogenous lysine amino acid excepting that when the endogenous amino acid in position X of said endogenous human GPCR is lysine, X_{codon} encodes an amino acid selected from the group consisting of histidine, arginine and alanine.
17. The nucleic acid sequence of claim 12 wherein X_{codon} is selected from the group consisting of AAA, AAG, GCA, GCG, GCC and GCU.
18. The nucleic acid sequence of claim 12 wherein X_{codon} is selected from the group consisting of AAA and AAG.
19. The nucleic acid sequence of claim 12 wherein P^{codon} is selected from the group consisting of CCA, CCC, CCG and CCU, and X_{codon} is selected from the group consisting of AAA and AAG.
20. A vector comprising the nucleic acid sequence of claim 12.
21. A plasmid comprising the nucleic acid sequence of claim 12.
22. A host cell comprising the nucleic acid sequence of claim 21.
23. The nucleic acid sequence of claim 12 in a purified and isolated form.
24. A method for selecting for alteration an endogenous amino acid residue within the third intracellular loop of a human G protein-coupled receptor ("GPCR"), said receptor

comprising a transmembrane 6 region and an intracellular loop 3 region, which endogenous amino acid, when altered to a non-endogenous amino acid, constitutively activates said human GPCR, comprising the following steps:

- 5 (a) identifying an endogenous proline residue within the transmembrane 6 region of a human GPCR;
- (b) identifying, by moving in a direction of the carboxy-terminus region of said GPCR towards the amino-terminus region of said GPCR, the endogenous, 16th amino acid residue from said proline residue;
- (c) altering the endogenous residue of step (b) to a non-endogenous amino acid
10 residue to create a non-endogenous version of an endogenous human GPCR; and
- (d) determining whether the non-endogenous human GPCR of step (c) is constitutively active.

25. The method of claim 24 wherein the amino acid residue that is two residues from said
15 proline residue in the transmembrane 6 region, in a carboxy-terminus to amino-terminus direction, is tryptophan.

26. A constitutively active, non-endogenous human GPCR produced by the process of claim 24.

27. A constitutively active, non-endogenous human GPCR produced by the process of
20 claim 25.

28. An algorithmic approach for creating a non-endogenous, constitutively active version of an endogenous human G protein coupled receptor (GPCR), said endogenous GPCR comprising a transmembrane 6 region and an intracellular loop 3 region, the

- 77 -

algorithmic approach comprising the steps of:

- (a) selecting an endogenous human GPCR comprising a proline residue in the transmembrane-6 region;
 - (b) identifying, by counting 16 amino acid residues from the proline residue of step (a), in a carboxy-terminus to amino-terminus direction, an endogenous amino acid residue;
 - (c) altering the identified amino acid residue of step (b) to a non-endogenous amino acid residue to create a non-endogenous version of the endogenous human GPCR; and
 - (d) determining if the non-endogenous version of the endogenous human GPCR of step (c) is constitutively active.
29. The algorithmic approach of claim 28 wherein the amino acid residue that is two residues from said proline residue in the transmembrane 6 region, in a carboxy-terminus to amino-terminus direction, is tryptophan.
30. A constitutively active, non-endogenous human GPCR produced by the algorithmic approach of claim 28.
31. A constitutively active, non-endogenous human GPCR produced by the algorithmic approach of claim 29.
32. A method for directly identifying a compound selected from the group consisting of inverse agonists, agonists and partial agonists to a non-endogenous, constitutively activated human G protein coupled receptor, said receptor comprising a transmembrane-6 region and an intracellular loop-3 region, comprising the steps of:
- (a) selecting an endogenous human GPCR;

- (b) identifying a proline residue within the transmembrane-6 region of the GPCR of step (a);
- (c) identifying, in a carboxy-terminus to amino-terminus direction, the endogenous, 16th amino acid residue from the proline residue of step (b);
- 5 (d) altering the endogenous amino acid of step (c) to a non-endogenous amino acid;
- (e) confirming that the non-endogenous GPCR of step (d) is constitutively active;
- (f) contacting a candidate compound with the non-endogenous, constitutively-activated GPCR of step (e); and
- 10 (g) determining, by measurement of the compound efficacy at said contacted receptor, whether said compound is an inverse agonist, agonist or partial agonist of said receptor.

- 33. The method of claim 32 wherein the non-endogenous amino acid of step (d) is lysine.
- 34. A compound directly identified by the method of claim 32.
- 15 35. The method of claim 32 wherein the directly identified compound is an inverse agonist.
- 36. The method of claim 32 wherein the directly identified compound is an agonist.--
- 37. The method of claim 32 wherein the directly identified compound is a partial agonist.
- 38. A composition comprising the inverse agonist of claim 35.
- 20 39. A composition comprising the agonist of claim 36.
- 40. A composition comprising the partial agonist of claim 37.
- 41. A method for directly identifying an inverse agonist to a non-endogenous,

- 79 -

constitutively activated human G protein coupled receptor ("GPCR"), said GPCR comprising a transmembrane-6 region and an intracellular loop-3 region, comprising the steps of:

- (a) selecting an endogenous human GPCR;
 - (b) identifying a proline residue within the transmembrane-6 region of the GPCR of
5 step (a);
 - (c) identifying, in a carboxy-terminus to amino-terminus direction, the
 endogenous, 16th amino acid residue from the proline residue of step (b);
 - (d) altering the endogenous amino acid of step (c) to a non-endogenous lysine residue;
 - (e) confirming that the non-endogenous GPCR of step (d) is constitutively active;
 - 10 (f) contacting a candidate compound with the non-endogenous, constitutively-
 activated GPCR of step (e); and
 - (g) determining, by measurement of the compound efficacy at said contacted receptor,
 whether said compound is an inverse agonist of said receptor.
42. An inverse agonist directly identified by the method of claim 37.
- 15 43. A composition comprising an inverse agonist of claim 38.

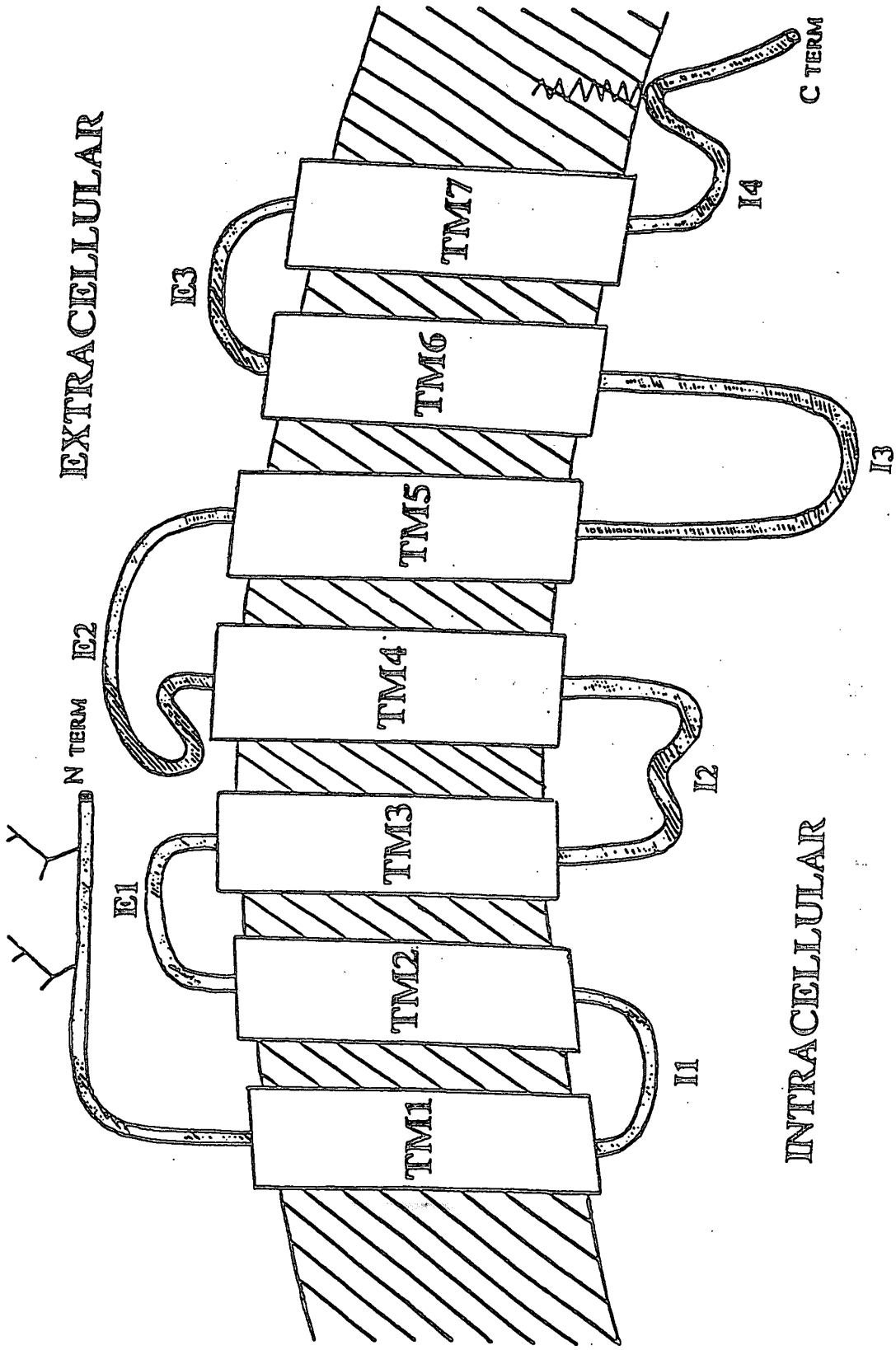


FIGURE 1

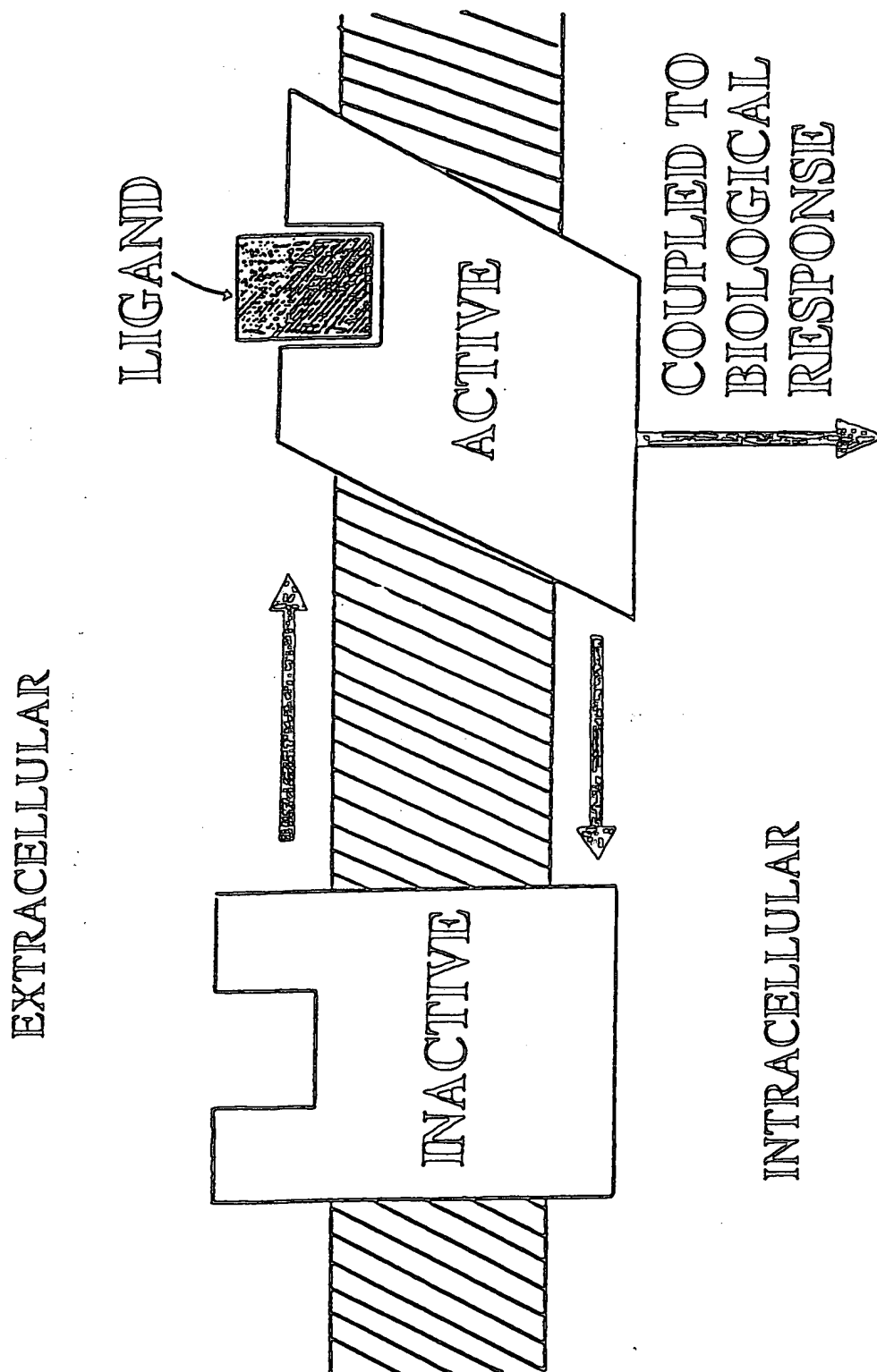


FIGURE 2

TTTGCTCAGAAGAAATGCCATCTAGTGATGATGAGGCTACTGCTGACTCTCAACATTCTACTCCTCCAAAAAAGAAGAGA 400
AAACGAGTCTTCTTTACGGTAGATCACTACTACTCCGATGACGACTGAGAGTTGTAAGATGAGGAGGTTTTTCTTCTCT

L L R R N A I G Y C . L S T F Y S S K K E E
F C S E E H P S S D D E A T A D S Q H S T P P K K K R
F A Q K K C H L V H H R L L L T L N I L L L O K R R E
K S L L F A H . H H H P . Q Q S E V N . E E L F S S F
Q E S S I G D L S S S A V A S E . C E V G G F F F L
K A . F F H W R T I I L S S S V R L H R S R W F L L S

Sty I

AAGGTAGAAGACCCCAAGGACTTTCTTCAGAATTGCTAAGTTTTTGAGTCATGCTGTGTTTAGTAATAGAACTCTTGC 480
TTCCATCTTCTGGGGTTCTTGAAAGGAAGTCTTAACGATTCAAAAACTCAGTACGACACAAATCATTATCTTGAGAACG

K G R R P Q G L S F R I A K F F E S C C V . . . N S C
K V E D P K D F P S E L L S F L S H A V F S N R T L A
R . K T P R T F L O N C . V F . V M L C L V I E L L
P L L G W P S E K L I A L N K S D H Q T . Y Y F E Q
F T S S G L S K G E S N S L K K L . A T N L L L V R A
L Y F V G L V K R . F O . T K Q T H S H K T I S S K S

TTGCTTTGCTATTTACACCACAAAGGAAAAAGCTGCACTGCTATACAAGAAAATTATGGAAAAATATTCTGTAACCTTTA 560
AACGAAACGATAAATGTGGTGTTCCTTTTTCGACGTGACGATATGTTCTTTTAATACCTTTTTATAAGACATTGGAAAT

L L C Y L H H K G K S C T A I O E N Y G K I F C N L Y
C F A I Y T T K E K A A L L Y K K I H E K Y S V T F
L A L L F T P O R K K L H C Y T R K L W K N I L . P L
K S Q . K C W L P F L O V A I C S F . P F I N O L R .
O K A I . V V F S F A A S S Y L F I I S F Y E T V K I
A K S N V G C L F F S C O . V L F N H F F I R Y G K

Ase I

TAAGTAGGCATAACAGTTATAATCATAACATACTGTTTTTTCTTACTCCACACAGGCATAGAGTGTCTGCTATTAATAAC 640
ATTCATCCGTATTGTCAATATTAGTATTGTATGACAAAAAAGAATGAGGTGTGTCCGTATCTCACAGACGATAATTATTG

K . A . O L . S . H T V F S Y S T O A . S V C Y . .
I S R H N S Y N H N I L F F L T P H R H R V S A I N N
V G I T V I I I T Y C F F L L H T G I E C L L L I T
L Y A Y C N Y D Y C V T K E . E V C A Y L T O . . Y S
L L C L L . L . L H S N K R V G C L C L T D A I L L
Y T P H V T I I H V Y Q K K K S W V P H S H R S N I V

Rsa I

TATGCTCAAAAATTGTGTACCTTTAGCTTTTAAATTTGTAAGGGGTTAATAAGGAATATTTGATGTATAGTGCTTGAC 720
ATACGAGTTTTTAACACATGGAAATCGAAAAATTAACATTTCCCAATTATTCTTATAAACTACATATCACGGAAGTG

L C S K I V Y L . L F N L . R G . . G I F D V . C L D
Y A Q K L C T F S F L I C K G V N K E Y L H Y S A L T
H L K N C V P L A F . F V K G L I R N I . C I V P .
H E F I T Y R . S K L K Y L P . Y P I N S T Y H R S
A . F N H V K L K K I O L P T L L S Y K I Y L A K V
I S L F O T G K A K . N T F P N I L F I O H I T G O S

FIGURE 3B

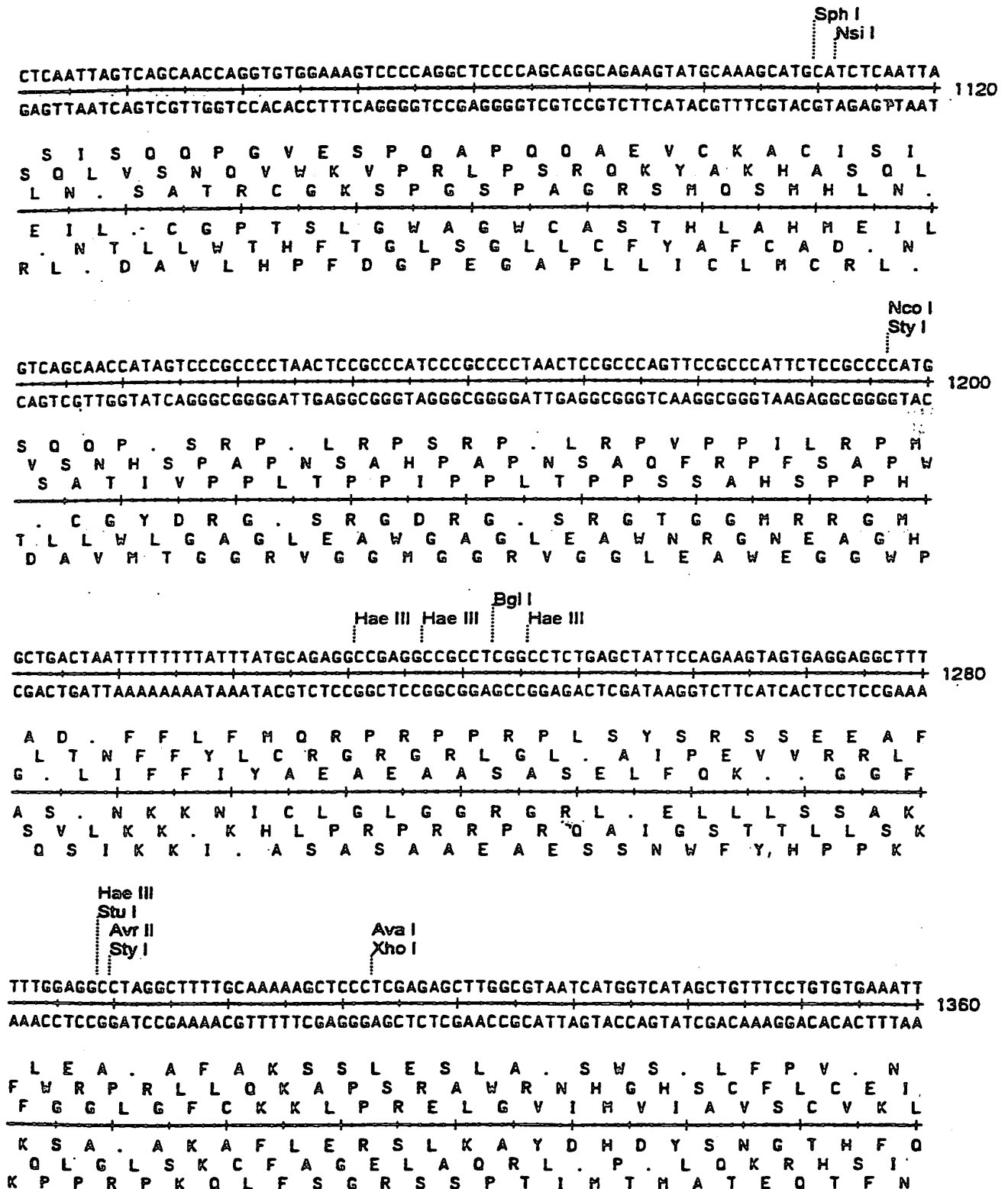


FIGURE 3D

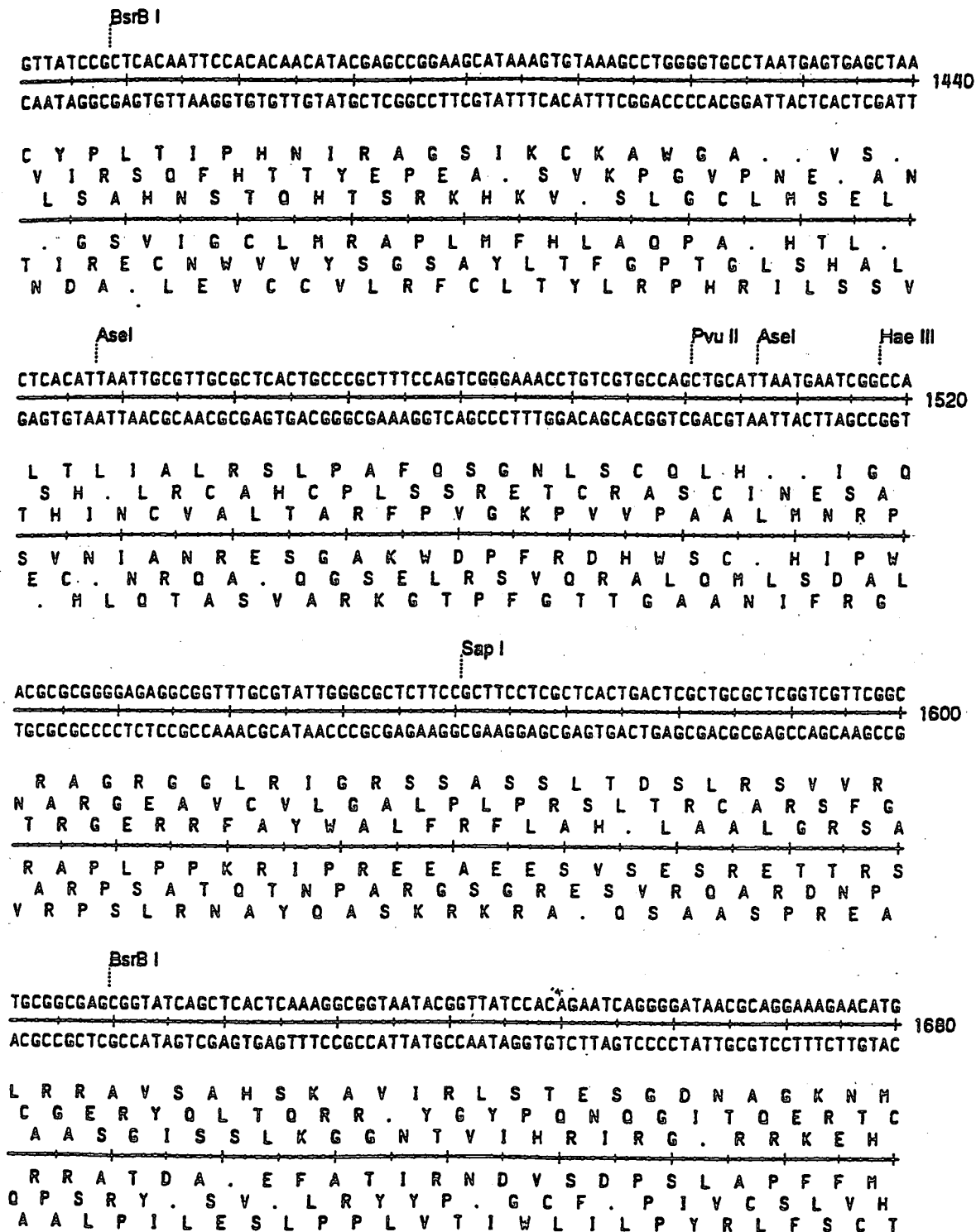


FIGURE 3E

Hae III Hae III Hae III

TGAGCAAAGGCCAGCAAAGGCCAGGAACCGTAAAAAGCCGCTTGCTGGCGTTTTTTCATAGGCTCCGCCCCCTGA 1760
ACTCGTTTTCCGGTCGTTTTCCGGTCCTTGGCATTITTCGGGCAACGACCGCAAAAAGGTATCCGAGGCGGGGGGACT

A K G O O K A R N R K K A A L L A F F H R L R P P D
E O K A S K R P G T V K R P R C W R F S I G S A P L
V S K R P A K G Q E P . K G R V A G V F P . A P P P .
H A F P W C F A L F R L F A A N S A N K W L S R G G S
S C F A L L L G P V T F L G R Q Q R K E M P E A G R V
L L L G A F P W S G Y F P R T A P T K G Y A G G G O

CGAGCATCACAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAGATACCAGGCGTTTCCCCCTG 1840
GCTCGTAGTGTTTTAGCTGCGAGTTCAGTCTCCACCGCTTTGGGCTGTCTGATATTTCTATGGTCCGCAAAGGGGGAC

E H H K N R R S S Q R W R N P T G L . R Y Q A F P P
T S I T K I D A Q V R G G E T R Q D Y K D T R R F P L
R A S O K S T L K S E V A K P D R T I K I P G V S P W
S C . L F R R E L . L H R F G V P S Y L Y W A N G G P
L M V F I S A . T L P P S V R C S . L S V L R K G R
R A D C F D V S L D S T A F G S L V I F I G P T E G O

GAAGCTCCCTCGTGCGCTCTCTGTTCCGACCTGCCGCTTACCGGATACCTGTCCGCTTTCTCCCTTCGGAAGCGTG 1920
CTTCGAGGGAGCACGCGAGAGGACAAGGCTGGGACGGCGAATGGCTATGGACAGGCGGAAAGAGGGAAGCCCTTCGCAC

G S S L V R S P V P T L P L T G Y L S A F L P S G S V
E A P S C A L L F R P C R L P D T C P P F S L R E A W
K L P R A L S C S D P A A Y R I P V R L S P F G K R
L E R T R E G T G V R G S V P Y R D A K R G E P L T
S A G E H A R R N R G O R K G S V Q G G K E R R S A H
F S G R A S E Q E S G A A . R I G T R R E G K P F R P

Apal I

GCGCTTTCTCAATGCTCAGCTGTAGGTATCTCAGTTCGGTGTAGGTGCTTCGCTCCAAGCTGGGCTGTGTGCACGAACC 2000
CGCGAAAGAGTTACGAGTGCGACATCCATAGAGTCAAGCCACATCCAGCAAGCGAGGTTGACCCGACACACGTGCTTGG

A L S Q C S R C R Y L S S V . V W R S K L G C V H E P
R F L N A H A V G I S V R C R S F A P S W A V C T N
G A F S M L T L . V S Q F G V G R S L Q A G L C A R T
A S E . H E R Q L Y R L E T Y T T R E L S P Q T C S G
R K R L A . A T P I E T R H L D N A G L O A T H V F G
A K E I S V S Y T D . N P T P R E S W A P S H A R V

Nci I

CCCCGTTACGCCCAGCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCAC 2080
GGGGCAAGTGGGCTGGCGACGCGGAATAGGCCATTGATAGCAGAACTCAGGTTGGGCCATTCTGTGCTGAATAGCGGTG

P V Q P D R C A L S G N Y R L E S N P V R H D L S P
P P F S P T A A P Y P V T I V L S P T R . D T T Y R H
P R S A R P L R L I R . L S S . V Q P G K T R L I A T
G T . G S R Q A K D P L . R R S D L G T L C S K D G S
G N L G V A A G . G T V I T K L G V R Y S V V . R W
G R E A R G S R R I R Y S D D Q T W G P L V R S I A V

FIGURE 3F

Ha III

TGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAAC 2160
ACCGTCGTCGGTGACCATTTGCTCTAATCGTCTCGCTCCATACATCCGCCACGATGTCTCAAGAACTTCACCACCGGATTG

L A A A T G N R I S R A R Y V G G A T E F L K W W P N
W O O P L V T G L A E R G M . A V L Q S S . S G G L T
G S S H W . Q D . Q S E V C R R C Y R V L E V V A .
A A A V P L L I L L A L Y T P P A V S N K F H H G L
O C C G S T V P N A S R P I Y A T S C L E Q L P P R V
P L L W Q Y C S . C L S T H L R H . L T R S T T A . S

TACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTC 2240
ATGCCGATGTGATCTTCTGTCTATAAACCATAGACGCGAGACGACTTCGGTCAATGGAAGCCTTTTCTCAACCATCGAG

Y G Y T R R T V F G I C A L L K P V T F G K R V G S S
T A T L E G O Y L V S A L C . S O L P S E K E L V A
L R L H . K D S I W Y L R S A E A S Y L R K K S W . L
P . V L L V T N P I Q A R S F G T V K P F L T P L E
V A V S S P C Y K T D A S Q Q L W N G E S F S N T A R
R S C . F S L I O Y R R E A S A L . R R F F L O Y S

TTGATCCGGCAAACAACCACCGCTGGTAGCGGTGGTTTTTTTGTGCAAGCAGCAGATTACGCGCAGAAAAAAGGAT 2320
AACTAGGCCGTTTTGTTGGTGGCGACCATCGCCACCAAAAAACAACGTTCTGTCGTCTAATGCGCGTCTTTTTTCTCTA

S G K Q T T A G S G G F F V C K Q Q I T R R K K G
L D P A N K P P L V A V V F L F A S S R L R A E K K D
L I R Q T N H R W . R W F F C L O A A D Y A Q K K R I
Q D P L C V V A P L P P K K T Q L C C I V R L F F P D
S G A F L G G S T A T T K K N A L L L N R A S F F S
K I R C V F W R Q Y R H N K Q K C A A S . A C F F L I

BspHI

CTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAAACGAAACTCACGTTAAGGGATTTTGGTCATG 2400
GAGTTCTTCTAGGAACTAGAAAAGATGCCCCAGACTGCGAGTCACCTTGCTTTTGAGTGCAATTCCTTAAAAACAGTAC

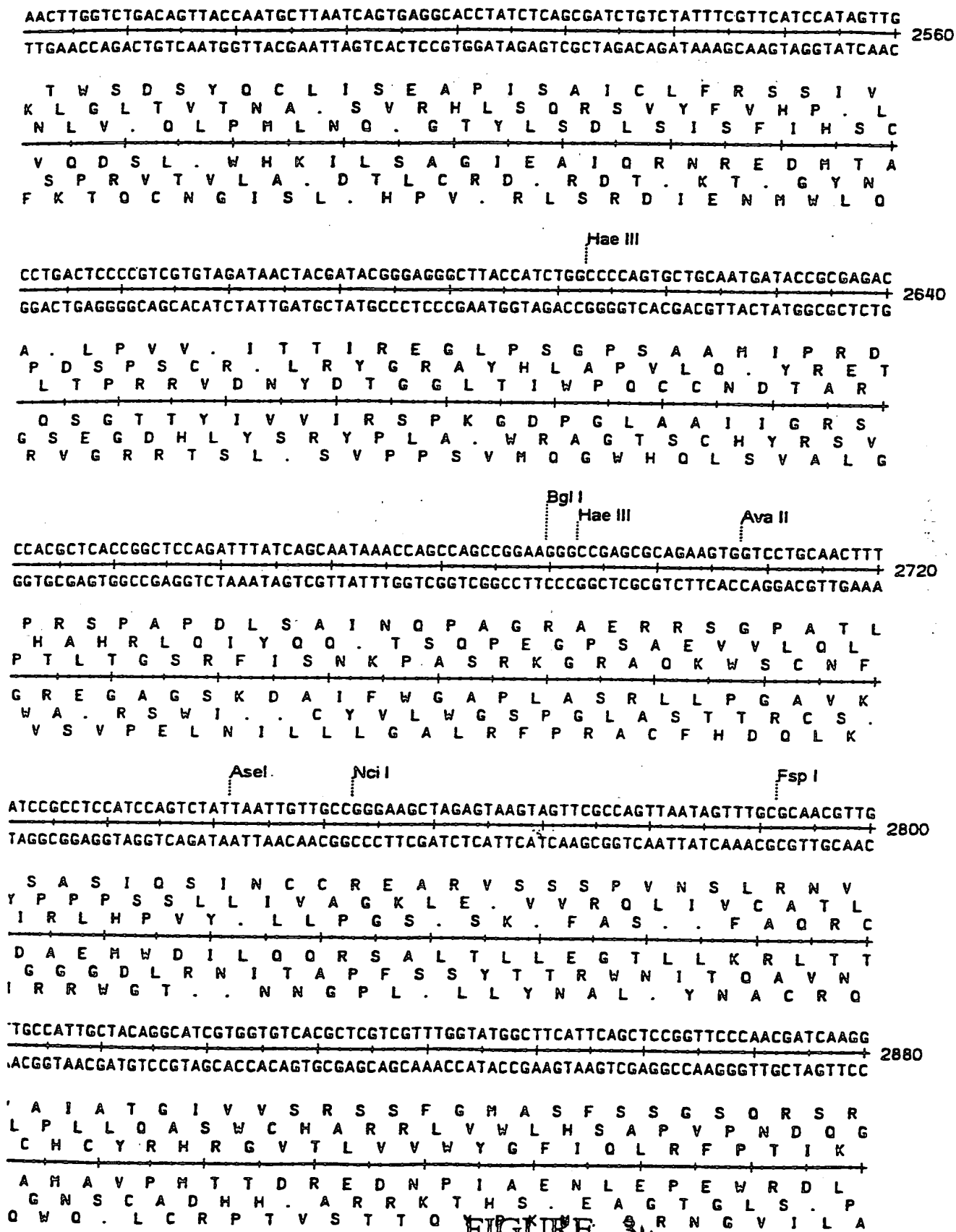
S O E D P L I F S T G S D A Q W N E N S R . G I L V H
L K K I L . S F L R G L T L S G T K T H V K G F W S .
S R R S F D L F Y G V . R S V E R K L T L R D F G H
S S G K I K E V P D S A . H F S F E R . P I K T H
R L F I R Q D K R R P R V S L P V F V . T L P N Q D H
E L L D K S R K . P T O R E T S R F S V N L S K P . S

DraI DraI

AGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAATAATGAAGTTTAAATCAATCTAAAGTATATATGAGTA 2480
TCTAATAGTTTTTCTAGAAGTGGATCTAGGAAAATTAATTTTACTTCAAAATTTAGTTAGATTTTCATATATACTCAT

R L S K R I F T . I L L N . K . S F K S I . S I Y E .
D Y O K G S S P R S F . I K N E V L N Q S K V Y H S
E I I K K D L H L D P F K L K M K F . I N L K Y I . V
L N D F L I K V . I R K F . F H L K L D I . L I Y S Y
S . . F P D E G L D K . I L F S T K F . D L T Y I L L
I I L F S R . R S G K L N F I F N . I L R F Y I H T

FIGURE 34



Ava II Pvu I Hae III

CGAGTTACATGATCCCCATGTTGTGCAAAAAGCGGTTAGCTCCTTCGGTCTCCGATCGTTGTCAGAAGTAAGTTGGC
 GCTCAATGTACTAGGGGGTACAACACGTTTTTTCGCCAATCGAGGAAGCCAGGAGGCTAGCAACAGTCTTCATTCAACCG 2960

R V T . S P H L C K K A V S S F G P P I V V R S K L A
 E L H D P P C C A K K R L A P S V L R S L S E V S W
 A S Y M I P H V V O K S G . L L R S S D R C O K . V G

R T V H D G M N H L F A T L E K P G G I T T L L L N A
 S N C S G G H O A F F R N A G E T R R D N D S T L O G
 L . M I G W T T C F L P . S R R D E S R Q . F Y T P

CGCAGTGTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGA
 GCGTCACAATAGTGAGTACCAATACCGTCGTGACGTATTAAGAGAATGACAGTACGGTAGGCATTCTACGAAAAGACACT 3040

A V L S L M V H A A L H N S L T V H P S V R C F S V
 P Q C Y H S W L W Q H C I I L L L S C H P . D A F L .
 R S V I T H G Y G S T A . F S Y C H A I R K H L F C D

A T N D S M T I A A S C L E R V T M G D T L H K E T V
 C H . . E H N H C C Q M I R K S D H W G Y S A K R H
 R L T I V . P . P L V A Y N E . Q . A M R L I S K O S

Rsa I Sca I Nci I Hinc II

CTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGCGTCAACACGGGAT
 GACCACTCATGAGTTGGTTCAGTAAGACTCTTATCACATACGCCGCTGGCTCAACGAGAACGGGCCGCGAGTTGTGCCCTA 3120

T G E Y S T K S F . E . C M R R P S C S C P A S T R D
 L V S T Q P S H S E N S V C G D R V A L A R R O H G I
 W . V L N O V I L R I V Y A T E L L L P G V N T G

P S Y E V L D N Q S Y H I R R G L O E Q G A D V R S
 S T L V . G L . E S F L T H P S R T A R A R R . C P I
 Q H T S L W T H R L I T Y A A V S N S K G P T L V P Y

Dra I Xmn I

AATACCGGCCACATAGCAGAACTTTAAAGTGCTCATCATTGGAACGTTCTTCGGGGCGAAACTCTCAAGGATCTT
 TTATGGCGCGGTGTATCGTCTTGAAATTTTCACGAGTAGTAACCTTTTGCAAGAAGCCCCGCTTTTGAGAGTTCCTAGAA 3200

N T A P H S R T L K V L I I G K R S S G R K L S R I L
 I P R H I A E L . K C S S L E N V L R G E N S O G S
 . Y R A T . Q N F K S A H H W K T F F G A K T L K D L

L V A G C L L V K F T S H M P F R E E P R F S E L I K
 I G R W H A S S . F H E D N S F T R R P S F E . P D .
 Y R A V Y C F K L L A . . Q F V N K P A F V R L S R

FIGURE 3I

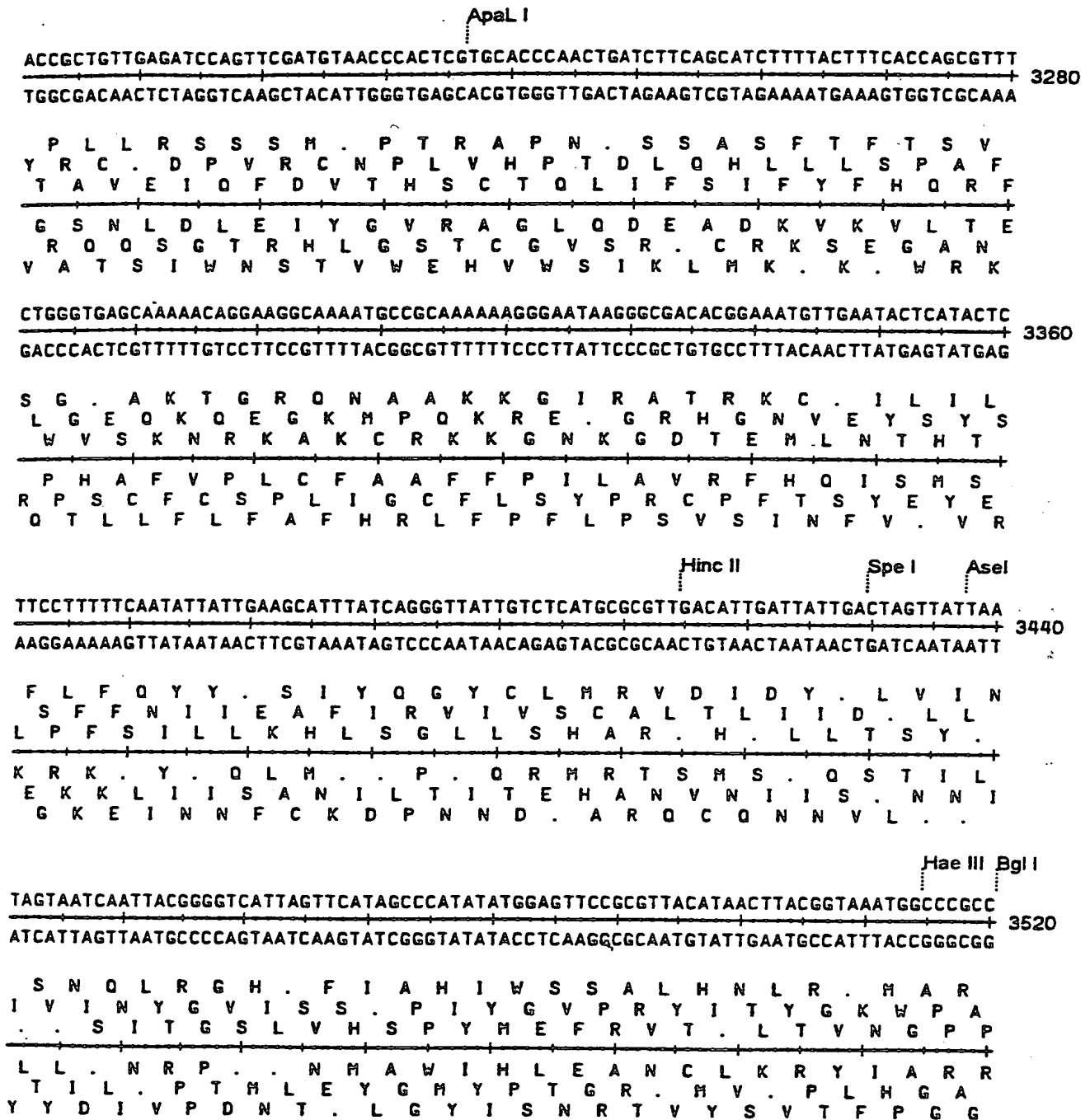


FIGURE 3J

Aat II

TGGCTGACCGCCCAACGACCCCCGCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCC
 ACCGACTGGCGGGTTGCTGGGGGCGGGTAACTGCAGTTATTACTGCATACAAGGGTATCATTGCGGTTATCCCTGAAAGG 3600

L A D R P T T P A H . R Q . . R H F P . . R Q . G L S
 W L T A Q R P P P I D V N N D V C S H S N A N R D F P
 G . P P N D P R P L T S I M T Y V P I V T P I G T F
 A S R G V V G A W Q R . Y H R I N G Y . Y R W Y P S E
 Q S V A W R G G G M S T L L S T H E W L L A L L S K G
 P O G G L S G R G N V D I I V Y T G H T V G I P V K W

Aat II Bgl I Rsa I Nde I Rsa I

ATTGACGTCAATGGGTGGACTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGATCATATGCCAAGTACGCCC
 TAACTGCAGTTACCCACCTGATAAATGCCATTTGACGGGTGAACCGTCATGTAGTTCACATAGTATACGGTTCATGCGGG 3680

I D V N G W T I Y G K L P T W Q Y I K C I I C O V R P
 L T S M G G L F T V N C P L G S T S S V S Y A K Y A
 H . R O W V D Y L R . T A H L A V H Q V Y H M P S T P
 H S T L P H V I . P L S G V O C Y M L H I M H W T R G
 N V D I P P S N V T F Q G S P L V D L T D Y A L Y A G
 O R . H T S . K R Y V A W K A T C . T Y . I G L V G

Aat II Hae III Bgl I Rsa I

CCTATTGACGTCAATGACGGTAAATGGCCCGCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCTACTTGGCA
 GGATAACTGCAGTTACTGCCATTTACCGGGCGGACCGTAATACGGGTCATGTACTGGAATACCCTGAAAGGATGAACCGT 3760

L L T S M T V N G P P G I M P S T . P Y G T F L L G
 P Y . R Q . R . M A R L A L C P V H D L M G L S Y L A
 P I D V N D G K W P A W H Y A Q Y M T L W D F P T W Q
 R N V D I V T F P G G P H I G L V H G . P V K R S P L
 . Q R . H R Y I A R R A N H G T C S R I P S E . K A
 G I S T L S P L H G A Q C . A W Y M V K H S K G V Q C

Rsa I BsaA I Nco I Rsa I
 SnaB I Sty I

GTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTG
 CATGTAGATGCATAATCAGTAGCGATAATGGTACCACTACGCCAAAACCGTCATGTAGTTACCCGCACCTATCGCCAAAC 3840

S T S T Y . S S L L P W . C G F G S T S M G V D S G L
 V H L R I S H R Y Y H G D A V L A V H Q W A W I A V .
 Y I Y V L V I A I T M V H R F W Q Y I N G R G . R F
 V D V Y . D D S N G H H H P K P L V D I P T S L P K
 I C R R I L . R . . W P S A T K A T C . H A H I A T Q
 Y M . T N T H A I V M T I R N Q C Y H L P R P Y R N S

FIGURE 3K

T H G D F Q V S T P L T S H G V C F G T K I N G T F Q
L T G I S K S P P H . R O W E F V L A P K S T G L S
D S R G F P S L H P I D V N G S L F W H Q N O R D F P
V . P S K W T E V G N V D I P T Q K P V L I L P V K W
S V P I E L D G G W Q R . H S N T K A G F D V P S E L
E R P N G L R W G H S T L P L K N Q C W F . R S K G

N V V T T P P H . R K W A V G V Y G G R S I . A E L S
K H S . Q L R P I D A N G R . A C T V G G R L Y K Q R S
K C R N N S A P L T Q H G G R R V R W E V Y I S Q R A L
F T T V V G G W Q R L H A T P T Y P P L D I Y A S S E
I D Y C S R G H S A F P R Y A H V T P P R Y L C L E
F H R L L E A G N V C I P P L R T R H S T . I L L A R

S G . L E N P L L L N W L I E I N T T H Y R E T
L A N . R T H C L T G L S K L I R L T I G R P
W L T R E P T A . L A Y R N . Y D S L . G D P
P . S S F G S S L Q S I S I L V V . L S V W
R A L . L V W Q K V P K D F N I R S V I P L G
Q S V L S G V A . S A . R F . Y S E S Y P S G

FIGURE 3L

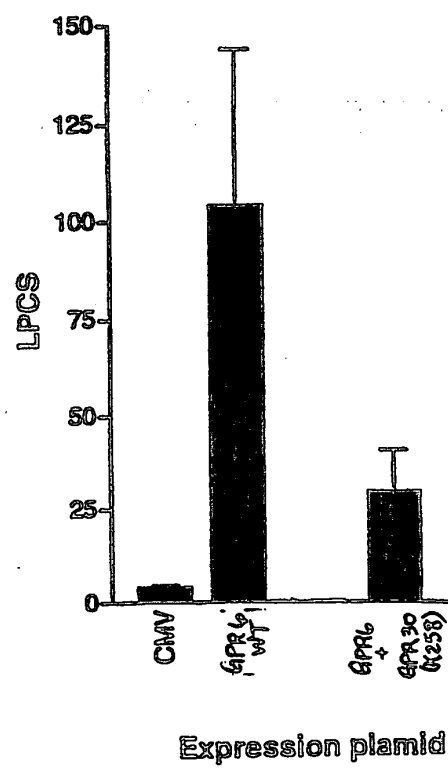


FIGURE 4

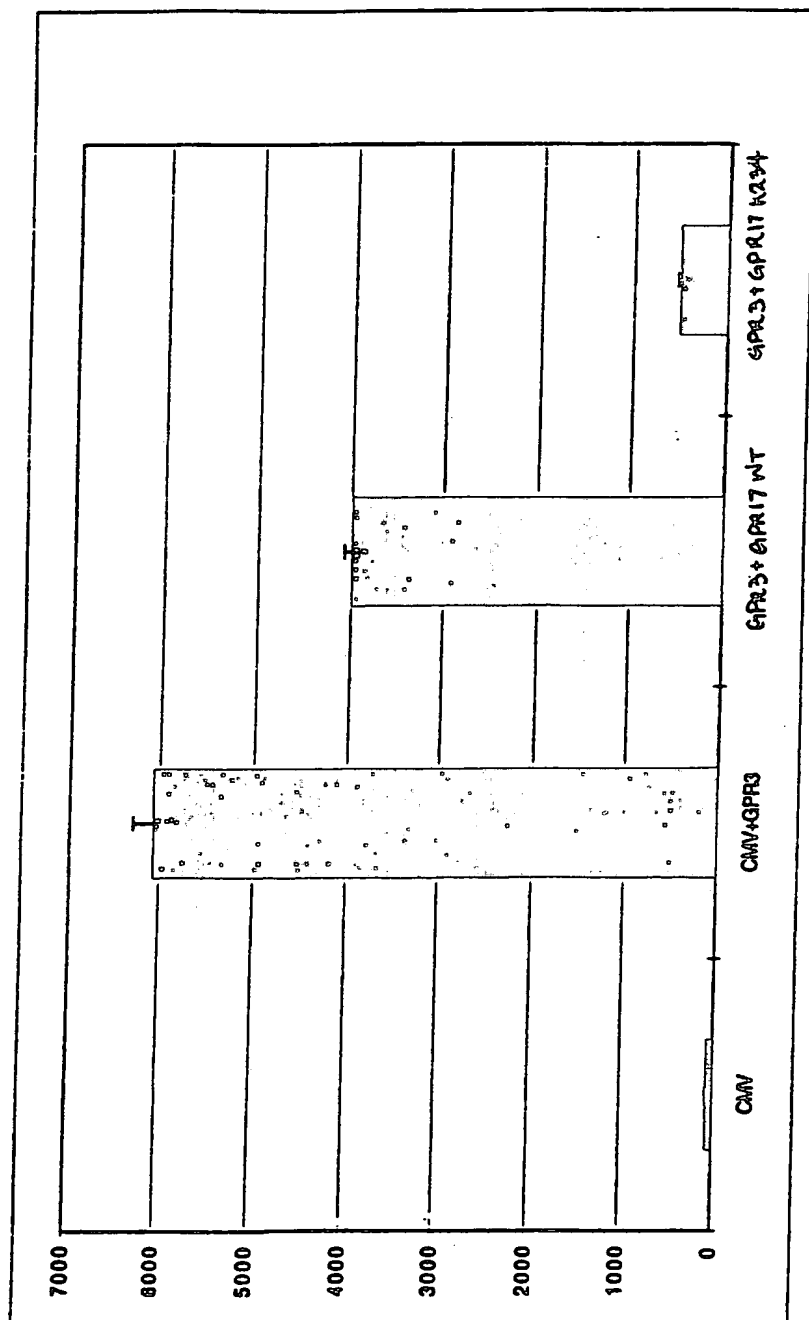


FIGURE 5

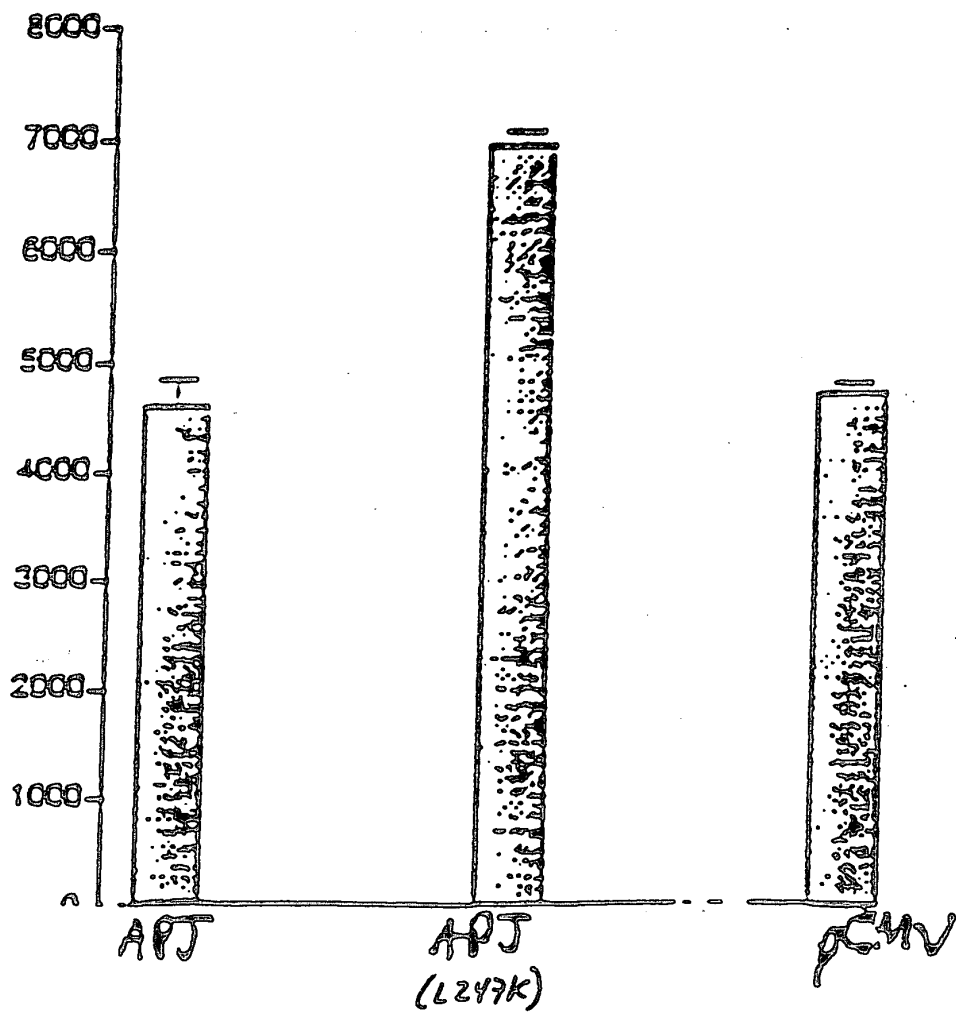


FIGURE 6

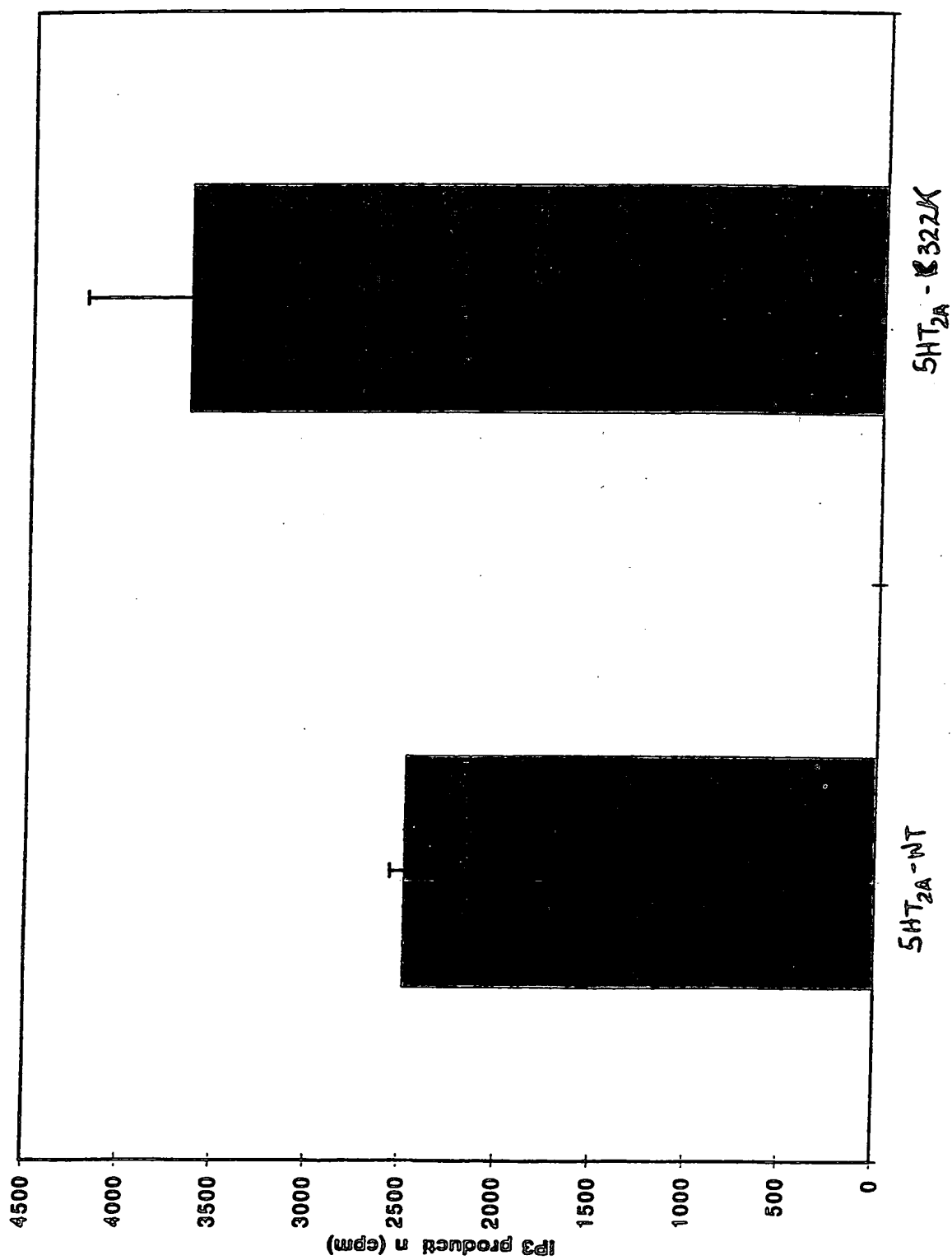


FIGURE 7

FIGURE 8A

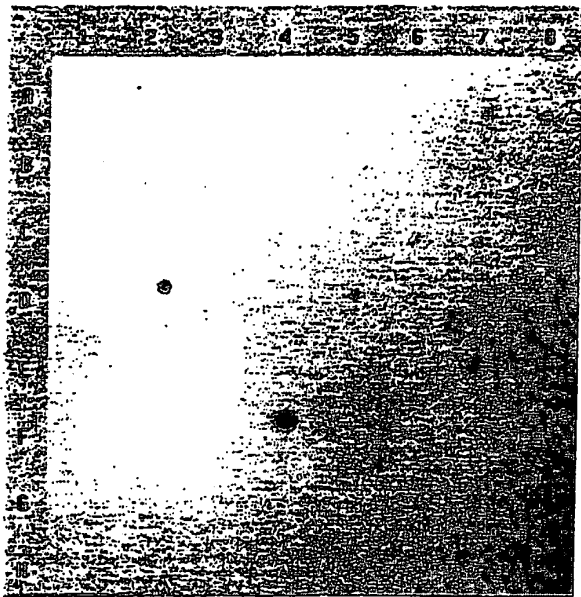


FIGURE 8B

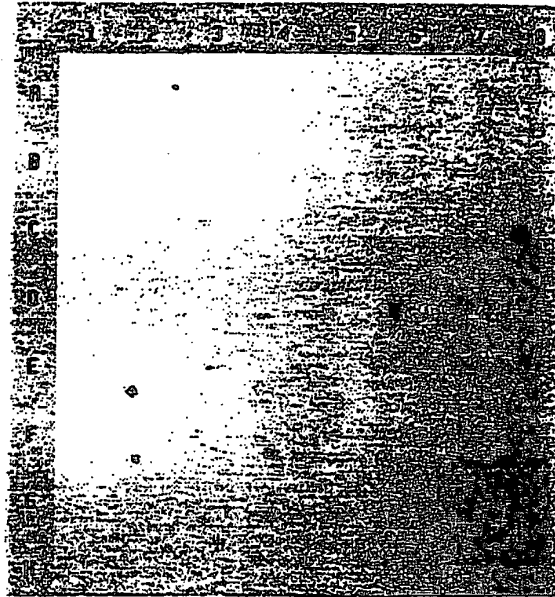


FIGURE 8C

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- 5 (i) APPLICANT: Behan, Dominic P.
Chalmers, Derek T.
Liaw, Chen W.
- (ii) TITLE OF INVENTION: Non-Endogenous, Constitutively
Activated Human G Protein-Coupled
Orphan Receptors
- 10 (iii) NUMBER OF SEQUENCES: 280
- (iv) CORRESPONDENCE ADDRESS:
15 (A) ADDRESSEE: Arena Pharmaceuticals, Inc.
(B) STREET: 6166 Nancy Ridge Drive
(C) CITY: San Diego
(D) STATE: CA
(E) COUNTRY: USA
(F) ZIP: 92122
- (v) COMPUTER READABLE FORM:
20 (A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
25 (A) APPLICATION NUMBER: US
(B) FILING DATE:
(C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Burgoon, Richard P.
(B) REGISTRATION NUMBER: 34,787
- 30 (ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (619)453-7200
(B) TELEFAX: (619)453-7210

(2) INFORMATION FOR SEQ ID NO:1:

- 35 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1068 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- 40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATGGAAGATT TGGAGGAAAC ATTATTTGAA GAATTTGAAA ACTATTCCTA TGACCTAGAC

60

TATTACTCTC TGGAGTCTGA TTTGGAGGAG AAAGTCCAGC TGGGAGTTGT TCACTGGGTC 120
 TCCCTGGTGT TATATTGTTT GGCTTTTGTT CTGGGAATTC CAGGAAATGC CATCGTCATT 180
 TGGTTCACGG GGCTCAAGTG GAAGAAGACA GTCACCACTC TGTGGTTCCT CAATCTAGCC 240
 ATTGCGGATT TCATTTTTCT TCTCTTTCTG CCCCTGTACA TCTCCTATGT GGCCATGAAT 300
 5 TTCCACTGGC CCTTTGGCAT CTGGCTGTGC AAAGCCAATT CCTTCACTGC CCAGTTGAAC 360
 ATGTTTGCCA GTGTTTTTTT CCTGACAGTG ATCAGCCTGG ACCACTATAT CCACTTGATC 420
 CATCCTGTCT TATCTCATCG GCATCGAACC CTCAAGAACT CTCTGATTGT CATTATATTC 480
 ATCTGGCTTT TGGCTTCTCT AATTGGCGGT CCTGCCCTGT ACTTCCGGGA CACTGTGGAG 540
 TTCAATAATC ATACTCTTTG CTATAACAAT TTTCAGAAGC ATGATCCTGA CCTCACTTTG 600
 10 ATCAGGCACC ATGTTCTGAC TTGGGTGAAA TTTATCATTG GCTATCTCTT CCCTTTGCTA 660
 ACAATGAGTA TTTGCTACTT GTGTCTCATC TTCAAGGTGA AGAAGCGAAC AGTCCTGATC 720
 TCCAGTAGGC ATTTCTGGAC AATTCTGGTT GTGGTTGTGG CCTTTGTGGT TTGCTGGACT 780
 CCTTATCACC TGTTTAGCAT TTGGGAGCTC ACCATTACACC ACAATAGCTA TTCCCACCAT 840
 GTGATGCAGG CTGGAATCCC CCTCTCCACT GGTTTGGCAT TCCTCAATAG TTGCTTGAAC 900
 15 CCCATCCTTT ATGTCCTAAT TAGTAAGAAG TTCCAAGCTC GCTTCCGGTC CTCAGTTGCT 960
 GAGATACTCA AGTACACACT GTGGGAAGTC AGCTGTTCTG GCACAGTGAG TGAACAGCTC 1020
 AGGAACTCAG AAACCAAGAA TCTGTGTCTC CTGGAAACAG CTCAATAA 1068

(3) INFORMATION FOR SEQ ID NO:2:

- 20 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 355 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

- 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Glu Asp Leu Glu Glu Thr Leu Phe Glu Glu Phe Glu Asn Tyr Ser
 1 5 10 15

Tyr Asp Leu Asp Tyr Tyr Ser Leu Glu Ser Asp Leu Glu Glu Lys Val
 20 25 30

30 Gln Leu Gly Val Val His Trp Val Ser Leu Val Leu Tyr Cys Leu Ala
 35 40 45

Phe Val Leu Gly Ile Pro Gly Asn Ala Ile Val Ile Trp Phe Thr Gly
 50 55 60

Leu Lys Trp Lys Lys Thr Val Thr Thr Leu Trp Phe Leu Asn Leu Ala
 65 70 75 80

5 Ile Ala Asp Phe Ile Phe Leu Leu Phe Leu Pro Leu Tyr Ile Ser Tyr
 85 90 95

Val Ala Met Asn Phe His Trp Pro Phe Gly Ile Trp Leu Cys Lys Ala
 100 105 110

10 Asn Ser Phe Thr Ala Gln Leu Asn Met Phe Ala Ser Val Phe Phe Leu
 115 120 125

Thr Val Ile Ser Leu Asp His Tyr Ile His Leu Ile His Pro Val Leu
 130 135 140

Ser His Arg His Arg Thr Leu Lys Asn Ser Leu Ile Val Ile Ile Phe
 145 150 155 160

15 Ile Trp Leu Leu Ala Ser Leu Ile Gly Gly Pro Ala Leu Tyr Phe Arg
 165 170 175

Asp Thr Val Glu Phe Asn Asn His Thr Leu Cys Tyr Asn Asn Phe Gln
 180 185 190

20 Lys His Asp Pro Asp Leu Thr Leu Ile Arg His His Val Leu Thr Trp
 195 200 205

Val Lys Phe Ile Ile Gly Tyr Leu Phe Pro Leu Leu Thr Met Ser Ile
 210 215 220

Cys Tyr Leu Cys Leu Ile Phe Lys Val Lys Lys Arg Thr Val Leu Ile
 225 230 235 240

25 Ser Ser Arg His Phe Trp Thr Ile Leu Val Val Val Ala Phe Val
 245 250 255

Val Cys Trp Thr Pro Tyr His Leu Phe Ser Ile Trp Glu Leu Thr Ile
 260 265 270

30 His His Asn Ser Tyr Ser His His Val Met Gln Ala Gly Ile Pro Leu
 275 280 285

Ser Thr Gly Leu Ala Phe Leu Asn Ser Cys Leu Asn Pro Ile Leu Tyr
 290 295 300

Val Leu Ile Ser Lys Lys Phe Gln Ala Arg Phe Arg Ser Ser Val Ala
 305 310 315 320

35 Glu Ile Leu Lys Tyr Thr Leu Trp Glu Val Ser Cys Ser Gly Thr Val
 325 330 335

Ser Glu Gln Leu Arg Asn Ser Glu Thr Lys Asn Leu Cys Leu Leu Glu

340

345

350

Thr Ala Gln
355

(4) INFORMATION FOR SEQ ID NO:3:

- 5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1089 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

- 10 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

	ATGGGCAACC ACACGTGGGA GGGCTGCCAC GTGGACTCGC GCGTGGACCA CCTCTTTC	60
	CCATCCCTCT ACATCTTTGT CATCGGCGTG GGGCTGCCCA CCAACTGCCT GGCTCTGTGG	120
	GCGGCCTACC GCCAGGTGCA ACAGCGCAAC GAGCTGGGCG TCTACCTGAT GAACCTCAGC	180
15	ATCGCCGACC TGCTGTACAT CTGCACGCTG CCGCTGTGGG TGGACTACTT CCTGCACCAC	240
	GACAACTGGA TCCACGGCCC CGGGTCCTGC AAGCTCTTTG GGTTCATCTT CTACACCAAT	300
	ATCTACATCA GCATCGCCTT CCTGTGCTGC ATCTCGGTGG ACCGCTACCT GGCTGTGGCC	360
	CACCCACTCC GCTTCGCCCC CCTGCGCCGC GTCAAGACCG CCGTGGCCGT GAGCTCCGTG	420
	GTCTGGGCCA CGGAGCTGGG CGCCAACTCG GCGCCCCTGT TCCATGACGA GCTCTTCCGA	480
20	GACCGCTACA ACCACACCTT CTGCTTTGAG AAGTTCCCCA TGGAAGGCTG GGTGGCCTGG	540
	ATGAACCTCT ATCGGGTGTT CGTGGGCTTC CTCTTCCCGT GGGCGCTCAT GCTGCTGTCG	600
	TACCGGGGCA TCCTGCGGGC CGTGCGGGGC AGCGTGTTCA CCGAGCGCCA GGAGAAGGCC	660
	AAGATCAAGC GGCTGGCCCT CAGCCTCATC GCCATCGTGC TGGTCTGCTT TGCGCCCTAT	720
	CACGTGCTCT TGCTGTCCCG CAGCGCCATC TACCTGGGCC GCCCCTGGGA CTGCGGCTTC	780
25	GAGGAGCGCG TCTTTTCTGC ATACCACAGC TCACTGGCTT TCACCAGCCT CAACTGTGTG	840
	GCGGACCCCA TCCTCTACTG CCTGGTCAAC GAGGGCGCCC GCAGCGATGT GGCCAAGGCC	900
	CTGCACAACC TGCTCCGCTT TCTGGCCAGC GACAAGCCCC AGGAGATGGC CAATGCCTCG	960
	CTCACCCTGG AGACCCCACT CACCTCCAAG AGGAACAGCA CAGCCAAAGC CATGACTGGC	1020
	AGCTGGGCGG CCACTCCGCC TTCCCAGGGG GACCAGGTGC AGCTGAAGAT GCTGCCGCCA	1080
30	GCACAATGA	1089

(5) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 362 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

10 Met Gly Asn His Thr Trp Glu Gly Cys His Val Asp Ser Arg Val Asp
 1 5 10 15
 His Leu Phe Pro Pro Ser Leu Tyr Ile Phe Val Ile Gly Val Gly Leu
 20 25 30
 Pro Thr Asn Cys Leu Ala Leu Trp Ala Ala Tyr Arg Gln Val Gln Gln
 35 40 45
 15 Arg Asn Glu Leu Gly Val Tyr Leu Met Asn Leu Ser Ile Ala Asp Leu
 50 55 60
 Leu Tyr Ile Cys Thr Leu Pro Leu Trp Val Asp Tyr Phe Leu His His
 65 70 75 80
 20 Asp Asn Trp Ile His Gly Pro Gly Ser Cys Lys Leu Phe Gly Phe Ile
 85 90 95
 Phe Tyr Thr Asn Ile Tyr Ile Ser Ile Ala Phe Leu Cys Cys Ile Ser
 100 105 110
 Val Asp Arg Tyr Leu Ala Val Ala His Pro Leu Arg Phe Ala Arg Leu
 115 120 125
 25 Arg Arg Val Lys Thr Ala Val Ala Val Ser Ser Val Val Trp Ala Thr
 130 135 140
 Glu Leu Gly Ala Asn Ser Ala Pro Leu Phe His Asp Glu Leu Phe Arg
 145 150 155 160
 30 Asp Arg Tyr Asn His Thr Phe Cys Phe Glu Lys Phe Pro Met Glu Gly
 165 170 175
 Trp Val Ala Trp Met Asn Leu Tyr Arg Val Phe Val Gly Phe Leu Phe
 180 185 190
 Pro Trp Ala Leu Met Leu Leu Ser Tyr Arg Gly Ile Leu Arg Ala Val
 195 200 205
 35 Arg Gly Ser Val Ser Thr Glu Arg Gln Glu Lys Ala Lys Ile Lys Arg
 210 215 220
 Leu Ala Leu Ser Leu Ile Ala Ile Val Leu Val Cys Phe Ala Pro Tyr

	225		230		235		240									
	His	Val	Leu	Leu	Leu	Ser	Arg	Ser	Ala	Ile	Tyr	Leu	Gly	Arg	Pro	Trp
					245					250					255	
5	Asp	Cys	Gly	Phe	Glu	Glu	Arg	Val	Phe	Ser	Ala	Tyr	His	Ser	Ser	Leu
			260						265					270		
	Ala	Phe	Thr	Ser	Leu	Asn	Cys	Val	Ala	Asp	Pro	Ile	Leu	Tyr	Cys	Leu
			275					280					285			
	Val	Asn	Glu	Gly	Ala	Arg	Ser	Asp	Val	Ala	Lys	Ala	Leu	His	Asn	Leu
		290					295					300				
10	Leu	Arg	Phe	Leu	Ala	Ser	Asp	Lys	Pro	Gln	Glu	Met	Ala	Asn	Ala	Ser
	305					310					315				320	
	Leu	Thr	Leu	Glu	Thr	Pro	Leu	Thr	Ser	Lys	Arg	Asn	Ser	Thr	Ala	Lys
					325					330					335	
15	Ala	Met	Thr	Gly	Ser	Trp	Ala	Ala	Thr	Pro	Pro	Ser	Gln	Gly	Asp	Gln
			340						345					350		
	Val	Gln	Leu	Lys	Met	Leu	Pro	Pro	Ala	Gln						
		355					360									

(6) INFORMATION FOR SEQ ID NO:5:

- 20 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

TATGAATTCA GATGCTCTAA ACGTCCCTGC

30

(7) INFORMATION FOR SEQ ID NO:6:

- 30 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

35 TCCGGATCCA CCTGCACCTG CGCCTGCACC

30

(8) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1002 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ATGGAGTCCT CAGGCAACCC AGAGAGCACC ACCTTTTTTTT ACTATGACCT TCAGAGCCAG 60
CCGTGTGAGA ACCAGGCCTG GGTCTTTGCT ACCCTCGCCA CCACTGTCCT GTACTGCCTG 120
10 GTGTTTCTCC TCAGCCTAGT GGGCAACAGC CTGGTCCTGT GGGTCCTGGT GAAGTATGAG 180
AGCCTGGAGT CCCTCACCAA CATCTTCATC CTCAACCTGT GCCTCTCAGA CCTGGTGTTT 240
GCCTGCTTGT TGCCTGTGTG GATCTCCCCA TACCACTGGG GCTGGGTGCT GGGAGACTTC 300
CTCTGCAAAC TCCTCAATAT GATCTTCTCC ATCAGCCTCT ACAGCAGCAT CTTCTTCCTG 360
ACCATCATGA CCATCCACCG CTACCTGTCG GTAGTGAGCC CCCTCTCCAC CCTGCGCGTC 420
15 CCCACCCTCC GCTGCCGGGT GCTGGTGACC ATGGCTGTGT GGGTAGCCAG CATCCTGTCC 480
TCCATCCTCG ACACCATCTT CCACAAGGTG CTTTCTTCGG GCTGTGATTA TTCCGAACTC 540
ACGTGGTACC TCACCTCCGT CTACCAGCAC AACCTCTTCT TCCTGCTGTC CCTGGGGATT 600
ATCCTGTTCT GCTACGTGGA GATCCTCAGG ACCCTGTTCC GCTCACGCTC CAAGCGGCGC 660
CACCGCACGG TCAAGCTCAT CTTGCCATC GTGGTGGCCT ACTTCCTCAG CTGGGGTCCC 720
20 TACAACTTCA CCCTGTTTCT GCAGACGCTG TTTCGGACCC AGATCATCCG GAGCTGCGAG 780
GCCAAACAGC AGCTAGAATA CGCCCTGCTC ATCTGCCGCA ACCTCGCCTT CTCCCACTGC 840
TGCTTTAACC CGGTGCTCTA TGTCTTCGTG GGGGTCAAGT TCCGCACACA CCTGAAACAT 900
GTTCTCCGGC AGTTCTGGTT CTGCCGGCTG CAGGCACCCA GCCCAGCCTC GATCCCCCAC 960
TCCCCTGGTG CCTTCGCCTA TGAGGGCGCC TCCTTCTACT GA 1002

25 (9) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 333 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

	Met	Glu	Ser	Ser	Gly	Asn	Pro	Glu	Ser	Thr	Thr	Phe	Phe	Tyr	Tyr	Asp	
	1				5					10						15	
	Leu	Gln	Ser	Gln	Pro	Cys	Glu	Asn	Gln	Ala	Trp	Val	Phe	Ala	Thr	Leu	
				20					25					30			
5	Ala	Thr	Thr	Val	Leu	Tyr	Cys	Leu	Val	Phe	Leu	Leu	Ser	Leu	Val	Gly	
			35					40					45				
	Asn	Ser	Leu	Val	Leu	Trp	Val	Leu	Val	Lys	Tyr	Glu	Ser	Leu	Glu	Ser	
		50					55					60					
10	Leu	Thr	Asn	Ile	Phe	Ile	Leu	Asn	Leu	Cys	Leu	Ser	Asp	Leu	Val	Phe	
	65				70						75					80	
	Ala	Cys	Leu	Leu	Pro	Val	Trp	Ile	Ser	Pro	Tyr	His	Trp	Gly	Trp	Val	
					85					90					95		
	Leu	Gly	Asp	Phe	Leu	Cys	Lys	Leu	Leu	Asn	Met	Ile	Phe	Ser	Ile	Ser	
				100					105						110		
15	Leu	Tyr	Ser	Ser	Ile	Phe	Phe	Leu	Thr	Ile	Met	Thr	Ile	His	Arg	Tyr	
			115						120					125			
	Leu	Ser	Val	Val	Ser	Pro	Leu	Ser	Thr	Leu	Arg	Val	Pro	Thr	Leu	Arg	
		130						135				140					
20	Cys	Arg	Val	Leu	Val	Thr	Met	Ala	Val	Trp	Val	Ala	Ser	Ile	Leu	Ser	
	145					150					155					160	
	Ser	Ile	Leu	Asp	Thr	Ile	Phe	His	Lys	Val	Leu	Ser	Ser	Gly	Cys	Asp	
				165						170					175		
	Tyr	Ser	Glu	Leu	Thr	Trp	Tyr	Leu	Thr	Ser	Val	Tyr	Gln	His	Asn	Leu	
				180					185					190			
25	Phe	Phe	Leu	Leu	Ser	Leu	Gly	Ile	Ile	Leu	Phe	Cys	Tyr	Val	Glu	Ile	
			195					200					205				
	Leu	Arg	Thr	Leu	Phe	Arg	Ser	Arg	Ser	Lys	Arg	Arg	His	Arg	Thr	Val	
		210						215				220					
30	Lys	Leu	Ile	Phe	Ala	Ile	Val	Val	Ala	Tyr	Phe	Leu	Ser	Trp	Gly	Pro	
	225					230					235					240	
	Tyr	Asn	Phe	Thr	Leu	Phe	Leu	Gln	Thr	Leu	Phe	Arg	Thr	Gln	Ile	Ile	
				245						250					255		
	Arg	Ser	Cys	Glu	Ala	Lys	Gln	Gln	Leu	Glu	Tyr	Ala	Leu	Leu	Ile	Cys	
				260					265					270			
35	Arg	Asn	Leu	Ala	Phe	Ser	His	Cys	Cys	Phe	Asn	Pro	Val	Leu	Tyr	Val	
			275					280					285				

Phe Val Gly Val Lys Phe Arg Thr His Leu Lys His Val Leu Arg Gln
 290 295 300

Phe Trp Phe Cys Arg Leu Gln Ala Pro Ser Pro Ala Ser Ile Pro His
 305 310 315 320

5 Ser Pro Gly Ala Phe Ala Tyr Glu Gly Ala Ser Phe Tyr
 325 330

(10) INFORMATION FOR SEQ ID NO:9:

- 10 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

15 GCAAGCTTGG GGGACGCCAG GTCGCCGGCT 30

(11) INFORMATION FOR SEQ ID NO:10:

- 20 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 31 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GCGGATCCGG ACGCTGGGGG AGTCAGGCTG C 31

25 (12) INFORMATION FOR SEQ ID NO:11:

- 30 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 987 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

ATGGACAACG CCTCGTTCTC GGAGCCCTGG CCCGCCAACG CATCGGGCCC GGACCCGGCG 60
 CTGAGCTGCT CCAACGCGTC GACTCTGGCG CCGCTGCCGG CGCCGCTGGC GGTGGCTGTA 120
 35 CCAGTTGTCT ACGCGGTGAT CTGCGCCGTG GGTCTGGCGG GCAACTCCGC CGTGCTGTAC 180

GTGTTGCTGC GGGCGCCCCG CATGAAGACC GTCACCAACC TGTTCATCCT CAACCTGGCC 240
 ATCGCCGACG AGCTCTTCAC GCTGGTGCTG CCCATCAACA TCGCCGACTT CCTGCTGCGG 300
 CAGTGGCCCT TCGGGGAGCT CATGTGCAAG CTCATCGTGG CTATCGACCA GTACAACACC 360
 TTCTCCAGCC TCTACTTCCT CACCGTCATG AGCGCCGACC GCTACCTGGT GGTGTTGGCC 420
 5 ACTGCGGAGT CGCGCCGGGT GGCCGGCCGC ACCTACAGCG CCGCGCGCGC GGTGAGCCTG 480
 GCCGTGTGGG GGATCGTCAC ACTCGTCGTG CTGCCCTTCG CAGTCTTCGC CCGGCTAGAC 540
 GACGAGCAGG GCCGGCGCCA GTGCGTGCTA GTCTTTCCGC AGCCCGAGGC CTTCTGGTGG 600
 CGCGCGAGCC GCCTCTACAC GTCGTGCTG GGCTTCGCCA TCCCCGTGTC CACCATCTGT 660
 GTCTCTATA CCACCCTGCT GTGCCGGCTG CATGCCATGC GGCTGGACAG CCACGCCAAG 720
 10 GCCCTGGAGC GCGCCAAGAA GCGGGTGACC TTCCTGGTGG TGGCAATCCT GGCGGTGTGC 780
 CTCCTCTGCT GGACGCCCTA CCACCTGAGC ACCGTGGTGG CGCTCACCAC CGACCTCCCG 840
 CAGACGCCGC TGGTCATCGC TATCTCCTAC TTCATCACCA GCCTGACGTA CGCCAACAGC 900
 TGCCTCAACC CCTTCCTCTA CGCCTTCCTG GACGCCAGCT TCCGCAGGAA CCTCCGCCAG 960
 CTGATAACTT GCCGCGCGGC AGCCTGA 987

15 (13) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 328 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 20 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Met Asp Asn Ala Ser Phe Ser Glu Pro Trp Pro Ala Asn Ala Ser Gly
 1 5 10 15
 25 Pro Asp Pro Ala Leu Ser Cys Ser Asn Ala Ser Thr Leu Ala Pro Leu
 20 25 30
 Pro Ala Pro Leu Ala Val Ala Val Pro Val Val Tyr Ala Val Ile Cys
 35 40 45
 30 Ala Val Gly Leu Ala Gly Asn Ser Ala Val Leu Tyr Val Leu Leu Arg
 50 55 60
 Ala Pro Arg Met Lys Thr Val Thr Asn Leu Phe Ile Leu Asn Leu Ala
 65 70 75 80

11

Ile Ala Asp Glu Leu Phe Thr Leu Val Leu Pro Ile Asn Ile Ala Asp
85 90 95

Phe Leu Leu Arg Gln Trp Pro Phe Gly Glu Leu Met Cys Lys Leu Ile
100 105 110

5 Val Ala Ile Asp Gln Tyr Asn Thr Phe Ser Ser Leu Tyr Phe Leu Thr
115 120 125

Val Met Ser Ala Asp Arg Tyr Leu Val Val Leu Ala Thr Ala Glu Ser
130 135 140

10 Arg Arg Val Ala Gly Arg Thr Tyr Ser Ala Ala Arg Ala Val Ser Leu
145 150 155 160

Ala Val Trp Gly Ile Val Thr Leu Val Val Leu Pro Phe Ala Val Phe
165 170 175

Ala Arg Leu Asp Asp Glu Gln Gly Arg Arg Gln Cys Val Leu Val Phe
180 185 190

15 Pro Gln Pro Glu Ala Phe Trp Trp Arg Ala Ser Arg Leu Tyr Thr Leu
195 200 205

Val Leu Gly Phe Ala Ile Pro Val Ser Thr Ile Cys Val Leu Tyr Thr
210 215 220

20 Thr Leu Leu Cys Arg Leu His Ala Met Arg Leu Asp Ser His Ala Lys
225 230 235 240

Ala Leu Glu Arg Ala Lys Lys Arg Val Thr Phe Leu Val Val Ala Ile
245 250 255

Leu Ala Val Cys Leu Leu Cys Trp Thr Pro Tyr His Leu Ser Thr Val
260 265 270

25 Val Ala Leu Thr Thr Asp Leu Pro Gln Thr Pro Leu Val Ile Ala Ile
275 280 285

Ser Tyr Phe Ile Thr Ser Leu Thr Tyr Ala Asn Ser Cys Leu Asn Pro
290 295 300

30 Phe Leu Tyr Ala Phe Leu Asp Ala Ser Phe Arg Arg Asn Leu Arg Gln
305 310 315 320

Leu Ile Thr Cys Arg Ala Ala Ala
325

(14) INFORMATION FOR SEQ ID NO:13:

- 35 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CGGAATTCGT CAACGGTCCC AGCTACAATG

30

(15) INFORMATION FOR SEQ ID NO:14:

5 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

ATGGATCCCA GGCCCTTCAG CACCGCAATA T

31

(16) INFORMATION FOR SEQ ID NO:15:

15 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1002 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

ATGCAGGCCG CTGGGCACCC AGAGCCCCTT GACAGCAGGG GTCCTTCTC CCTCCCCACG 60

ATGGGTGCCA ACGTCTCTCA GGACAATGGC ACTGGCCACA ATGCCACCTT CTCCGAGCCA 120

CTGCCGTTCC TCTATGTGCT CCTGCCCGCC GTGTACTCCG GGATCTGTGC TGTGGGGCTG 180

ACTGGCAACA CGGCCGTCAT CTTGTAATC CTAAGGGCGC CCAAGATGAA GACGGTGACC 240

25 AACGTGTTCA TCCTGAACCT GGCCGTCGCC GACGGGCTCT TCACGCTGGT ACTGCCCGTC 300

AACATCGCGG AGCACCTGCT GCAGTACTGG CCCTTCGGGG AGCTGCTCTG CAAGCTGGTG 360

CTGGCCGTCG ACCACTACAA CATCTTCTCC AGCATCTACT TCCTAGCCGT GATGAGCGTG 420

GACCGATACC TGGTGGTGCT GGCCACCGTG AGGTCCCGCC ACATGCCCTG GCGCACCTAC 480

CGGGGGGCGA AGGTCGCCAG CCTGTGTGTC TGGCTGGGCG TCACGGTCCT GGTTCGTGCC 540

30 TTCTTCTCTT TCGCTGGCGT CTACAGCAAC GAGCTGCAGG TCCCAAGCTG TGGGCTGAGC 600

TTCCCGTGGC CCGAGCGGGT CTGGTTCAAG GCCAGCCGTG TCTACACTTT GGTCTGGGC 660

TTCGTGCTGC CCGTGTGCAC CATCTGTGTG CTCTACACAG ACCTCCTGCG CAGGCTGCGG 720

GCCGTGCGGC TCCGCTCTGG AGCCAAGGCT CTAGGCAAGG CCAGGCGGAA GGTGACCGTC 780
 CTGGTCCTCG TCGTGCTGGC CGTGTGCCTC CTCTGCTGGA CGCCCTTCCA CCTGGCCTCT 840
 GTCGTGGCCC TGACCACGSA CCTGCCCCAG ACCCCACTGG TCATCAGTAT GTCCTACGTC 900
 ATCACCAGCC TCACGTACGC CAACTCGTGC CTGAACCCCT TCCTCTACGC CTTTCTAGAT 960
 5 GACAACTTCC GGAAGAACTT CCGCAGCATA TTGCGGTGCT GA 1002

(17) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 333 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

15 Met Gln Ala Ala Gly His Pro Glu Pro Leu Asp Ser Arg Gly Ser Phe
 1 5 10 15
 Ser Leu Pro Thr Met Gly Ala Asn Val Ser Gln Asp Asn Gly Thr Gly
 20 25 30
 His Asn Ala Thr Phe Ser Glu Pro Leu Pro Phe Leu Tyr Val Leu Leu
 35 40 45
 20 Pro Ala Val Tyr Ser Gly Ile Cys Ala Val Gly Leu Thr Gly Asn Thr
 50 55 60
 Ala Val Ile Leu Val Ile Leu Arg Ala Pro Lys Met Lys Thr Val Thr
 65 70 75 80
 25 Asn Val Phe Ile Leu Asn Leu Ala Val Ala Asp Gly Leu Phe Thr Leu
 85 90 95
 Val Leu Pro Val Asn Ile Ala Glu His Leu Leu Gln Tyr Trp Pro Phe
 100 105 110
 Gly Glu Leu Leu Cys Lys Leu Val Leu Ala Val Asp His Tyr Asn Ile
 115 120 125
 30 Phe Ser Ser Ile Tyr Phe Leu Ala Val Met Ser Val Asp Arg Tyr Leu
 130 135 140
 Val Val Leu Ala Thr Val Arg Ser Arg His Met Pro Trp Arg Thr Tyr
 145 150 155 160
 35 Arg Gly Ala Lys Val Ala Ser Leu Cys Val Trp Leu Gly Val Thr Val
 165 170 175

Leu Val Leu Pro Phe Phe Ser Phe Ala Gly Val Tyr Ser Asn Glu Leu
 180 185 190
 Gln Val Pro Ser Cys Gly Leu Ser Phe Pro Trp Pro Glu Arg Val Trp
 195 200 205
 5 Phe Lys Ala Ser Arg Val Tyr Thr Leu Val Leu Gly Phe Val Leu Pro
 210 215 220
 Val Cys Thr Ile Cys Val Leu Tyr Thr Asp Leu Leu Arg Arg Leu Arg
 225 230 235 240
 10 Ala Val Arg Leu Arg Ser Gly Ala Lys Ala Leu Gly Lys Ala Arg Arg
 245 250 255
 Lys Val Thr Val Leu Val Leu Val Val Leu Ala Val Cys Leu Leu Cys
 260 265 270
 Trp Thr Pro Phe His Leu Ala Ser Val Val Ala Leu Thr Thr Asp Leu
 275 280 285
 15 Pro Gln Thr Pro Leu Val Ile Ser Met Ser Tyr Val Ile Thr Ser Leu
 290 295 300
 Thr Tyr Ala Asn Ser Cys Leu Asn Pro Phe Leu Tyr Ala Phe Leu Asp
 305 310 315 320
 20 Asp Asn Phe Arg Lys Asn Phe Arg Ser Ile Leu Arg Cys
 325 330

(18) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 48 base pairs
 (B) TYPE: nucleic acid
 25 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: DNA (genomic)
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

ACGAATTCAG CCATGGTCCT TGAGGTGAGT GACCACCAAG TGCTAAAT

48

30 (19) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 27 base pairs
 (B) TYPE: nucleic acid
 35 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: DNA (genomic)
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GAGGATCCTG GAATGCGGGG AAGTCAG

27

(20) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

5

(A) LENGTH: 1107 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

10	ATGGTCCTTG AGGTGAGTGA CCACCAAGTG CTAAATGACG CCGAGGTTGC CGCCCTCCTG	60
	GAGAACTTCA GCTCTTCCTA TGA CTATGGA GAAAACGAGA GTGACTCGTG CTGTACCTCC	120
	CCGCCCTGCC CACAGGACTT CAGCCTGAAC TTCGACCGGG CCTTCCTGCC AGCCCTCTAC	180
	AGCCTCCTCT TTCTGCTGGG GCTGCTGGGC AACGGCGCGG TGGCAGCCGT GCTGCTGAGC	240
	CGGCGGACAG CCCTGAGCAG CACCGACACC TTCCTGCTCC ACCTAGCTGT AGCAGACACG	300
15	CTGCTGGTGC TGACACTGCC GCTCTGGGCA GTGGACGCTG CCGTCCAGTG GGTCTTTGGC	360
	TCTGGCCTCT GCAAAGTGGC AGGTGCCCTC TTCAACATCA ACTTCTACGC AGGAGCCCTC	420
	CTGCTGGCCT GCATCAGCTT TGACCGCTAC CTGAACATAG TTCATGCCAC CCAGCTCTAC	480
	CGCCGGGGGC CCCC GGCCCG CGTGACCCTC ACCTGCCTGG CTGTCTGGGG GCTCTGCCTG	540
	CTTTTCGCCC TCCCAGACTT CATCTTCCTG TCGGCCCACC ACGACGAGCG CCTCAACGCC	600
20	ACCCACTGCC AATACAACTT CCCACAGGTG GGCCGCACGG CTCTGCGGGT GCTGCAGCTG	660
	GTGGCTGGCT TTCTGCTGCC CCTGCTGGTC ATGGCCTACT GCTATGCCCA CATCCTGGCC	720
	GTGCTGCTGG TTTCCAGGGG CCAGCGGCGC CTGCGGGCCA TCGGCTGGT GGTGGTGGTC	780
	GTGGTGGCCT TTGCCCTCTG CTGGACCCCT TATCACCTGG TGGTGCTGGT GGACATCCTC	840
	ATGGACCTGG GCGCTTTGGC CCGCAACTGT GGCCGAGAAA GCAGGGTAGA CGTGGCCAAG	900
25	TCGGTCACCT CAGGCCTGGG CTACATGCAC TGCTGCCTCA ACCCGCTGCT CTATGCCTTT	960
	G TAGGGGTCA AGTTCCGGGA GCGGATGTGG ATGCTGCTCT TGCGCCTGGG CTGCCCCAAC	1020
	CAGAGAGGGC TCCAGAGGCA GCCATCGTCT TCCCGCCGGG ATTCATCCTG GTCTGAGACC	1080
	TCAGAGGCCT CCTACTCGGG CTTGTGA	1107

(21) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 368 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Met Val Leu Glu Val Ser Asp His Gln Val Leu Asn Asp Ala Glu Val
 1 5 10 15
 Ala Ala Leu Leu Glu Asn Phe Ser Ser Ser Tyr Asp Tyr Gly Glu Asn
 10 20 25 30
 Glu Ser Asp Ser Cys Cys Thr Ser Pro Pro Cys Pro Gln Asp Phe Ser
 35 40 45
 Leu Asn Phe Asp Arg Ala Phe Leu Pro Ala Leu Tyr Ser Leu Leu Phe
 15 50 55 60
 Leu Leu Gly Leu Leu Gly Asn Gly Ala Val Ala Ala Val Leu Leu Ser
 65 70 75 80
 Arg Arg Thr Ala Leu Ser Ser Thr Asp Thr Phe Leu Leu His Leu Ala
 85 90 95
 Val Ala Asp Thr Leu Leu Val Leu Thr Leu Pro Leu Trp Ala Val Asp
 100 105 110
 Ala Ala Val Gln Trp Val Phe Gly Ser Gly Leu Cys Lys Val Ala Gly
 115 120 125
 Ala Leu Phe Asn Ile Asn Phe Tyr Ala Gly Ala Leu Leu Leu Ala Cys
 125 130 135 140
 Ile Ser Phe Asp Arg Tyr Leu Asn Ile Val His Ala Thr Gln Leu Tyr
 145 150 155 160
 Arg Arg Gly Pro Pro Ala Arg Val Thr Leu Thr Cys Leu Ala Val Trp
 165 170 175
 Gly Leu Cys Leu Leu Phe Ala Leu Pro Asp Phe Ile Phe Leu Ser Ala
 180 185 190
 His His Asp Glu Arg Leu Asn Ala Thr His Cys Gln Tyr Asn Phe Pro
 195 200 205
 Gln Val Gly Arg Thr Ala Leu Arg Val Leu Gln Leu Val Ala Gly Phe
 210 215 220
 Leu Leu Pro Leu Leu Val Met Ala Tyr Cys Tyr Ala His Ile Leu Ala
 225 230 235 240

17

Val Leu Leu Val Ser Arg Gly Gln Arg Arg Leu Arg Ala Met Arg Leu
 245 250 255

Val Val Val Val Val Val Ala Phe Ala Leu Cys Trp Thr Pro Tyr His
 260 265 270

5 Leu Val Val Leu Val Asp Ile Leu Met Asp Leu Gly Ala Leu Ala Arg
 275 280 285

Asn Cys Gly Arg Glu Ser Arg Val Asp Val Ala Lys Ser Val Thr Ser
 290 295 300

10 Gly Leu Gly Tyr Met His Cys Cys Leu Asn Pro Leu Leu Tyr Ala Phe
 305 310 315 320

Val Gly Val Lys Phe Arg Glu Arg Met Trp Met Leu Leu Leu Arg Leu
 325 330 335

Gly Cys Pro Asn Gln Arg Gly Leu Gln Arg Gln Pro Ser Ser Ser Arg
 340 345 350

15 Arg Asp Ser Ser Trp Ser Glu Thr Ser Glu Ala Ser Tyr Ser Gly Leu
 355 360 365

(22) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 20 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

25 TTAAGCTTGA CCTAATGCCA TCTTGTGTCC 30

(23) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 30 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

TTGGATCCAA AAGAACCATG CACCTCAGAG 30

35 (24) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1074 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

```

ATGGCTGATG ACTATGGCTC TGAATCCACA TCTTCCATGG AAGACTACGT TAACTTCAAC    60
TTCCTGACT TCTACTGTGA GAAAAACAAT GTCAGGCAGT TTGCGAGCCA TTCCTCCCA    120
CCCTTGTA CTGGCTCGTGT CATCGTGGGT GCCTTGGGCA ACAGTCTTGT TATCCTTGTC    180
10 TACTGGTACT GCACAAGAGT GAAGACCATG ACCGACATGT TCCTTTTGAA TTTGGCAATT    240
GCTGACCTCC TCTTCTTGT CACTCTTCCC TTCTGGGCCA TTGCTGCTGC TGACCAGTGG    300
AAGTTCAG CTTTCATGTG CAAGGTGGTC AACAGCATGT ACAAGATGAA CTTCTACAGC    360
TGTGTGTTGC TGATCATGTG CATCAGCGTG GACAGGTACA TTGCCATTGC CCAGGCCATG    420
AGAGCACATA CTTGGAGGGA GAAAAGGCTT TTGTACAGCA AAATGGTTTG CTTTACCATC    480
15 TGGGTATTGG CAGCTGCTCT CTGCATCCCA GAAATCTTAT ACAGCCAAAT CAAGGAGGAA    540
TCCGGCATTG CTATCTGCAC CATGGTTTAC CCTAGCGATG AGAGCACCAA ACTGAAGTCA    600
GCTGTCTTGA CCCTGAAGGT CATTCTGGGG TTCTTCCTTC CCTTCGTGGT CATGGCTTGC    660
TGCTATACCA TCATCATTCA CACCCTGATA CAAGCCAAGA AGTCTTCCA GCACAAAGCC    720
CTAAAAGTGA CCATCACTGT CCTGACCGTC TTTGTCTTGT CTCAGTTTCC CTACAACTGC    780
20 ATTTTGTGG TGCAGACCAT TGACGCCTAT GCCATGTTCA TCTCCAAGT TGCCGTTTCC    840
ACCAACATTG ACATCTGCTT CCAGGTCACC CAGACCATCG CTTCTTCCA CAGTTGCCTG    900
AACCCTGTT TCTATGTTTT TGTGGGTGAG AGATTCCGCC GGGATCTCGT GAAAACCCTG    960
AAGAACTTGG GTTGCATCAG CCAGGCCAG TGGGTTTCAT TTACAAGGAG AGAGGGAAGC   1020
TTGAAGCTGT CGTCTATGTT GCTGGAGACA ACCTCAGGAG CACTCTCCCT CTGA       1074

```

25 (25) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 357 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant

30

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

	Met	Ala	Asp	Asp	Tyr	Gly	Ser	Glu	Ser	Thr	Ser	Ser	Met	Glu	Asp	Tyr	
	1				5					10					15		
5	Val	Asn	Phe	Asn	Phe	Thr	Asp	Phe	Tyr	Cys	Glu	Lys	Asn	Asn	Val	Arg	
				20					25					30			
	Gln	Phe	Ala	Ser	His	Phe	Leu	Pro	Pro	Leu	Tyr	Trp	Leu	Val	Phe	Ile	
			35					40					45				
	Val	Gly	Ala	Leu	Gly	Asn	Ser	Leu	Val	Ile	Leu	Val	Tyr	Trp	Tyr	Cys	
		50					55					60					
10	Thr	Arg	Val	Lys	Thr	Met	Thr	Asp	Met	Phe	Leu	Leu	Asn	Leu	Ala	Ile	
	65					70					75					80	
	Ala	Asp	Leu	Leu	Phe	Leu	Val	Thr	Leu	Pro	Phe	Trp	Ala	Ile	Ala	Ala	
					85					90					95		
15	Ala	Asp	Gln	Trp	Lys	Phe	Gln	Thr	Phe	Met	Cys	Lys	Val	Val	Asn	Ser	
				100					105					110			
	Met	Tyr	Lys	Met	Asn	Phe	Tyr	Ser	Cys	Val	Leu	Leu	Ile	Met	Cys	Ile	
			115					120					125				
	Ser	Val	Asp	Arg	Tyr	Ile	Ala	Ile	Ala	Gln	Ala	Met	Arg	Ala	His	Thr	
		130					135					140					
20	Trp	Arg	Glu	Lys	Arg	Leu	Leu	Tyr	Ser	Lys	Met	Val	Cys	Phe	Thr	Ile	
	145					150					155					160	
	Trp	Val	Leu	Ala	Ala	Ala	Leu	Cys	Ile	Pro	Glu	Ile	Leu	Tyr	Ser	Gln	
				165						170					175		
25	Ile	Lys	Glu	Glu	Ser	Gly	Ile	Ala	Ile	Cys	Thr	Met	Val	Tyr	Pro	Ser	
				180					185					190			
	Asp	Glu	Ser	Thr	Lys	Leu	Lys	Ser	Ala	Val	Leu	Thr	Leu	Lys	Val	Ile	
			195					200					205				
	Leu	Gly	Phe	Phe	Leu	Pro	Phe	Val	Val	Met	Ala	Cys	Cys	Tyr	Thr	Ile	
		210					215					220					
30	Ile	Ile	His	Thr	Leu	Ile	Gln	Ala	Lys	Lys	Ser	Ser	Lys	His	Lys	Ala	
	225					230					235					240	
	Leu	Lys	Val	Thr	Ile	Thr	Val	Leu	Thr	Val	Phe	Val	Leu	Ser	Gln	Phe	
				245					250						255		
35	Pro	Tyr	Asn	Cys	Ile	Leu	Leu	Val	Gln	Thr	Ile	Asp	Ala	Tyr	Ala	Met	
				260					265					270			
	Phe	Ile	Ser	Asn	Cys	Ala	Val	Ser	Thr	Asn	Ile	Asp	Ile	Cys	Phe	Gln	
			275					280					285				

Val Thr Gln Thr Ile Ala Phe Phe His Ser Cys Leu Asn Pro Val Leu
 290 295 300

Tyr Val Phe Val Gly Glu Arg Phe Arg Arg Asp Leu Val Lys Thr Leu
 305 310 315 320

5 Lys Asn Leu Gly Cys Ile Ser Gln Ala Gln Trp Val Ser Phe Thr Arg
 325 330 335

Arg Glu Gly Ser Leu Lys Leu Ser Ser Met Leu Leu Glu Thr Thr Ser
 340 345 350

10 Gly Ala Leu Ser Leu
 355

(26) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1110 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

ATGGCCTCAT CGACCACTCG GGGCCCCAGG GTTCTGACT TATTTTCTGG GCTGCCGCCG 60

20 GCGGTCACAA CTCCCGCCAA CCAGAGCGCA GAGGCCTCGG CGGGCAACGG GTCGGTGGCT 120

GGCGCGGACG CTCCAGCCGT CACGCCCTTC CAGAGCCTGC AGCTGGTGCA TCAGCTGAAG 180

GGGCTGATCG TGCTGCTCTA CAGCGTCGTG GTGGTCGTGG GGCTGGTGGG CAACTGCCTG 240

CTGGTGCTGG TGATCGCGCG GGTGCCGCGG CTGCACAACG TGACGAACTT CCTCATCGGC 300

AACCTGGCCT TGTCCGACGT GTCATGTGC ACCGCCTGCG TGCCGCTCAC GCTGGCCTAT 360

25 GCCTTCGAGC CACGCGGCTG GGTGTTCCGC GGC GGCCCTGT GCCACCTGGT CTTCTTCCTG 420

CAGCCGGTCA CCGTCTATGT GTCGGTGTTT ACGCTCACCA CCATCGCAGT GGACCGCTAC 480

GTCGTGCTGG TGCACCCGCT GAGGCGCGCA TCTCGCTGCG CCTCAGCCTA CGCTGTGCTG 540

GCCATCTGGG CGCTGTCCGC GGTGCTGGCG CTGCCGCCCG CCGTGACAC CTATCACGTG 600

GAGCTCAAGC CGCACGACGT GCGCCTCTGC GAGGAGTTCT GGGGCTCCCA GGAGCGCCAG 660

30 CGCCAGCTCT ACGCCTGGGG GCTGCTGCTG GTCACCTACC TGCTCCCTCT GCTGGTCATC 720

CTCCTGTCTT ACGTCCGGGT GTCAGTGAAG CTCCGCAACC GCGTGGTGCC GGGCTGCGTG 780

ACCCAGAGCC AGGCCGACTG GGACCGCGCT CGGCGCCGGC GCACCTTCTG CTTGCTGGTG 840

GTGGTCGTGG TGGTGTTCGC CGTCTGCTGG CTGCCGCTGC ACGTCTTCAA CCTGCTGCGG 900
 GACCTCGACC CCCACGCCAT CGACCCTTAC GCCTTTGGGC TGGTGCAGCT GCTCTGCCAC 960
 TGGCTCGCCA TGAGTTCGGC CTGCTACAAC CCCTTCATCT ACGCTTGGCT GCACGACAGC 1020
 TTCCGCGAGG AGCTGCGCAA ACTGTTGGTC GCTTGGCCCC GCAAGATAGC CCCCCATGGC 1080
 5 CAGAATATGA CCGTCAGCGT GGTCATCTGA 1110

(27) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 369 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

15 Met Ala Ser Ser Thr Thr Arg Gly Pro Arg Val Ser Asp Leu Phe Ser
 1 5 10 15
 Gly Leu Pro Pro Ala Val Thr Thr Pro Ala Asn Gln Ser Ala Glu Ala
 20 20 25 30
 Ser Ala Gly Asn Gly Ser Val Ala Gly Ala Asp Ala Pro Ala Val Thr
 35 40 45
 20 Pro Phe Gln Ser Leu Gln Leu Val His Gln Leu Lys Gly Leu Ile Val
 50 55 60
 Leu Leu Tyr Ser Val Val Val Val Gly Leu Val Gly Asn Cys Leu
 65 70 75 80
 25 Leu Val Leu Val Ile Ala Arg Val Pro Arg Leu His Asn Val Thr Asn
 85 90 95
 Phe Leu Ile Gly Asn Leu Ala Leu Ser Asp Val Leu Met Cys Thr Ala
 100 105 110
 Cys Val Pro Leu Thr Leu Ala Tyr Ala Phe Glu Pro Arg Gly Trp Val
 115 120 125
 30 Phe Gly Gly Gly Leu Cys His Leu Val Phe Phe Leu Gln Pro Val Thr
 130 135 140
 Val Tyr Val Ser Val Phe Thr Leu Thr Thr Ile Ala Val Asp Arg Tyr
 145 150 155 160
 35 Val Val Leu Val His Pro Leu Arg Arg Ala Ser Arg Cys Ala Ser Ala
 165 170 175

Tyr Ala Val Leu Ala Ile Trp Ala Leu Ser Ala Val Leu Ala Leu Pro
 180 185 190
 Pro Ala Val His Thr Tyr His Val Glu Leu Lys Pro His Asp Val Arg
 195 200 205
 5 Leu Cys Glu Glu Phe Trp Gly Ser Gln Glu Arg Gln Arg Gln Leu Tyr
 210 215 220
 Ala Trp Gly Leu Leu Leu Val Thr Tyr Leu Leu Pro Leu Leu Val Ile
 225 230 235 240
 10 Leu Leu Ser Tyr Val Arg Val Ser Val Lys Leu Arg Asn Arg Val Val
 245 250 255
 Pro Gly Cys Val Thr Gln Ser Gln Ala Asp Trp Asp Arg Ala Arg Arg
 260 265 270
 Arg Arg Thr Phe Cys Leu Leu Val Val Val Val Val Val Phe Ala Val
 275 280 285
 15 Cys Trp Leu Pro Leu His Val Phe Asn Leu Leu Arg Asp Leu Asp Pro
 290 295 300
 His Ala Ile Asp Pro Tyr Ala Phe Gly Leu Val Gln Leu Leu Cys His
 305 310 315 320
 20 Trp Leu Ala Met Ser Ser Ala Cys Tyr Asn Pro Phe Ile Tyr Ala Trp
 325 330 335
 Leu His Asp Ser Phe Arg Glu Glu Leu Arg Lys Leu Leu Val Ala Trp
 340 345 350
 Pro Arg Lys Ile Ala Pro His Gly Gln Asn Met Thr Val Ser Val Val
 355 360 365
 25 Ile

(28) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1083 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 30
 (ii) MOLECULE TYPE: DNA (genomic)
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:
 35 ATGGACCCAG AAGAACTTC AGTTTATTTG GATTATTACT ATGCTACGAG CCCAACTCT 60
 GACATCAGGG AGACCCACTC CCATGTTCTT TACACCTCTG TCTTCCTTCC AGTCTTTTAC 120

ACAGCTGTGT TCCTGACTGG AGTGCTGGGG AACCTTGTTT TCATGGGAGC GTTGCATTTT 180
 AAACCCGGCA GCCGAAGACT GATCGACATC TTTATCATCA ATCTGGCTGC CTCTGACTTC 240
 ATTTTTCTTG TCACATTGCC TCTCTGGGTG GATAAAGAAG CATCTCTAGG ACTGTGGAGG 300
 ACGGGCTCCT TCCTGTGCAA AGGGAGCTCC TACATGATCT CCGTCAATAT GCACTGCAGT 360
 5 GTCCTCCTGC TCACTTGCAT GAGTGTGAC CGCTACCTGG CCATTGTGTG GCCAGTCGTA 420
 TCCAGGAAAT TCAGAAGGAC AGACTGTGCA TATGTAGTCT GTGCCAGCAT CTGGTTTATC 480
 TCCTGCCTGC TGGGGTTGCC TACTCTTCTG TCCAGGGAGC TCACGCTGAT TGATGATAAG 540
 CCATACTGTG CAGAGAAAAA GGCAACTCCA ATTAACTCA TATGGTCCCT GGTGGCCTTA 600
 ATTTTCACCT TTTTGTCCC TTTGTTGAGC ATTGTGACCT GCTACTGTTG CATTGCAAGG 660
 10 AAGCTGTGTG CCCATTACCA GCAATCAGGA AAGCACAACA AAAAGCTGAA GAAATCTATA 720
 AAGATCATCT TTATTGTCGT GGCAGCCTTT CTTGTCTCCT GGCTGCCCTT CAATACTTTC 780
 AAGTTCCTGG CCATTGTCTC TGGGTTGCGG CAAGAACACT ATTTACCCTC AGCTATTCTT 840
 CAGCTTGGTA TGGAGGTGAG TGGACCCTTG GCATTTGCCA ACAGCTGTGT CAACCCTTTC 900
 ATTTACTATA TCTTCGACAG CTACATCCGC CGGGCCATTG TCCACTGCTT GTGCCCTTGC 960
 15 CTGAAAAACT ATGACTTTGG GAGTAGCACT GAGACATCAG ATAGTCACCT CACTAAGGCT 1020
 CTCTCCACCT TCATTCATGC AGAAGATTTT GCCAGGAGGA GGAAGAGGTC TGTGTCACTC 1080
 TAA 1083

(29) INFORMATION FOR SEQ ID NO:28:

20 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 360 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

Met Asp Pro Glu Glu Thr Ser Val Tyr Leu Asp Tyr Tyr Tyr Ala Thr
 1 5 10 15
 Ser Pro Asn Ser Asp Ile Arg Glu Thr His Ser His Val Pro Tyr Thr
 20 25 30
 30 Ser Val Phe Leu Pro Val Phe Tyr Thr Ala Val Phe Leu Thr Gly Val
 35 40 45

Leu Gly Asn Leu Val Leu Met Gly Ala Leu His Phe Lys Pro Gly Ser
 50 55 60
 Arg Arg Leu Ile Asp Ile Phe Ile Ile Asn Leu Ala Ala Ser Asp Phe
 65 70 75 80
 5 Ile Phe Leu Val Thr Leu Pro Leu Trp Val Asp Lys Glu Ala Ser Leu
 85 90 95
 Gly Leu Trp Arg Thr Gly Ser Phe Leu Cys Lys Gly Ser Ser Tyr Met
 100 105 110
 10 Ile Ser Val Asn Met His Cys Ser Val Leu Leu Leu Thr Cys Met Ser
 115 120 125
 Val Asp Arg Tyr Leu Ala Ile Val Trp Pro Val Val Ser Arg Lys Phe
 130 135 140
 Arg Arg Thr Asp Cys Ala Tyr Val Val Cys Ala Ser Ile Trp Phe Ile
 145 150 155 160
 15 Ser Cys Leu Leu Gly Leu Pro Thr Leu Leu Ser Arg Glu Leu Thr Leu
 165 170 175
 Ile Asp Asp Lys Pro Tyr Cys Ala Glu Lys Lys Ala Thr Pro Ile Lys
 180 185 190
 20 Leu Ile Trp Ser Leu Val Ala Leu Ile Phe Thr Phe Phe Val Pro Leu
 195 200 205
 Leu Ser Ile Val Thr Cys Tyr Cys Cys Ile Ala Arg Lys Leu Cys Ala
 210 215 220
 His Tyr Gln Gln Ser Gly Lys His Asn Lys Lys Leu Lys Lys Ser Ile
 225 230 235 240
 25 Lys Ile Ile Phe Ile Val Val Ala Ala Phe Leu Val Ser Trp Leu Pro
 245 250 255
 Phe Asn Thr Phe Lys Phe Leu Ala Ile Val Ser Gly Leu Arg Gln Glu
 260 265 270
 30 His Tyr Leu Pro Ser Ala Ile Leu Gln Leu Gly Met Glu Val Ser Gly
 275 280 285
 Pro Leu Ala Phe Ala Asn Ser Cys Val Asn Pro Phe Ile Tyr Tyr Ile
 290 295 300
 Phe Asp Ser Tyr Ile Arg Arg Ala Ile Val His Cys Leu Cys Pro Cys
 305 310 315 320
 35 Leu Lys Asn Tyr Asp Phe Gly Ser Ser Thr Glu Thr Ser Asp Ser His
 325 330 335
 Leu Thr Lys Ala Leu Ser Thr Phe Ile His Ala Glu Asp Phe Ala Arg

340

345

350

Arg Arg Lys Arg Ser Val Ser Leu
355 360

(30) INFORMATION FOR SEQ ID NO:29:

- 5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

- 10 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

CTAGAATTCT GACTCCAGCC AAAGCATGAA T

31

(31) INFORMATION FOR SEQ ID NO:30:

- 15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GCTGGATCCT AACAGTCTG CGCTCGGCCT

30

(32) INFORMATION FOR SEQ ID NO:31:

- 25 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1020 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

30 ATGAATGGCC TTGAAGTGGC TCCCCAGGT CTGATCACCA ACTTCTCCCT GGCCACGGCA 60
GAGCAATGTG GCCAGGAGAC GCCACTGGAG AACATGCTGT TCGCCTCCTT CTACCTTCTG 120
GATTTTATCC TGGCTTTAGT TGGCAATACC CTGGCTCTGT GGCTTTTCAT CCGAGACCAC 180
AAGTCCGGGA CCCC GGCCAA CGTGTTCTTG ATGCATCTGG CCGTGGCCGA CTTGTCTGTC 240
GTGCTGGTCC TGCCCACCCG CCTGGTCTAC CACTTCTCTG GGAACCACTG GCCATTTGGG 300

GAAATCGCAT GCCGTCTCAC CGGCTTCCTC TTCTACCTCA ACATGTACGC CAGCATCTAC 360
 TTCCTCACCT GCATCAGCGC CGACCGTTTC CTGGCCATTG TGCACCCGGT CAAGTCCCTC 420
 AAGCTCCGCA GGGCCCTCTA CGCACACCTG GCCTGTGCCT TCCTGTGGGT GGTGGTGGCT 480
 GTGGCCATGG CCCCCTGCT GGTGAGCCCA CAGACCGTGC AGACCAACCA CACGGTGGTC 540
 5 TGCCTGCAGC TGTACCGGGA GAAGGCCTCC CACCATGCCC TGGTGTCCCT GGCAGTGGCC 600
 TTCACCTTCC CGTTCATCAC CACGGTCACC TGCTACCTGC TGATCATCCG CAGCCTGCGG 660
 CAGGGCCTGC GTGTGGAGAA GCGCCTCAAG ACCAAGGCAG TGCGCATGAT CGCCATAGTG 720
 CTGGCCATCT TCCTGGTCTG CTTCTGCCCC TACCACGTCA ACCGCTCCGT CTACGTGCTG 780
 CACTACCGCA GCCATGGGGC CTCCTGCGCC ACCCAGCGCA TCCTGGCCCT GGCAAACCGC 840
 10 ATCACCTCCT GCCTCACCAG CCTCAACGGG GCACTCGACC CCATCATGTA TTTCTTCGTG 900
 GCTGAGAAGT TCCGCCACGC CCTGTGCAAC TTGCTCTGTG GCAAAGGCT CAAGGGCCCG 960
 CCCCCAGCT TCGAAGGGAA AACCAACGAG AGCTCGCTGA GTGCCAAGTC AGAGCTGTGA 1020

(33) INFORMATION FOR SEQ ID NO:32:

15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 339 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

Met Asn Gly Leu Glu Val Ala Pro Pro Gly Leu Ile Thr Asn Phe Ser
 1 5 10 15

Leu Ala Thr Ala Glu Gln Cys Gly Gln Glu Thr Pro Leu Glu Asn Met
 20 25 30

25 Leu Phe Ala Ser Phe Tyr Leu Leu Asp Phe Ile Leu Ala Leu Val Gly
 35 40 45

Asn Thr Leu Ala Leu Trp Leu Phe Ile Arg Asp His Lys Ser Gly Thr
 50 55 60

30 Pro Ala Asn Val Phe Leu Met His Leu Ala Val Ala Asp Leu Ser Cys
 65 70 75 80

Val Leu Val Leu Pro Thr Arg Leu Val Tyr His Phe Ser Gly Asn His
 85 90 95

Trp Pro Phe Gly Glu Ile Ala Cys Arg Leu Thr Gly Phe Leu Phe Tyr

27

		100			105			110
		Leu Asn Met Tyr Ala Ser Ile Tyr Phe Leu Thr Cys Ile Ser Ala Asp						
		115			120			125
5		Arg Phe Leu Ala Ile Val His Pro Val Lys Ser Leu Lys Leu Arg Arg						
		130		135			140	
		Pro Leu Tyr Ala His Leu Ala Cys Ala Phe Leu Trp Val Val Val Ala						
		145		150		155		160
		Val Ala Met Ala Pro Leu Leu Val Ser Pro Gln Thr Val Gln Thr Asn						
			165		170			175
10		His Thr Val Val Cys Leu Gln Leu Tyr Arg Glu Lys Ala Ser His His						
			180		185			190
		Ala Leu Val Ser Leu Ala Val Ala Phe Thr Phe Pro Phe Ile Thr Thr						
			195		200			205
		Val Thr Cys Tyr Leu Leu Ile Ile Arg Ser Leu Arg Gln Gly Leu Arg						
15			210		215			220
		Val Glu Lys Arg Leu Lys Thr Lys Ala Val Arg Met Ile Ala Ile Val						
			225		230		235	240
		Leu Ala Ile Phe Leu Val Cys Phe Val Pro Tyr His Val Asn Arg Ser						
			245		250			255
20		Val Tyr Val Leu His Tyr Arg Ser His Gly Ala Ser Cys Ala Thr Gln						
			260		265			270
		Arg Ile Leu Ala Leu Ala Asn Arg Ile Thr Ser Cys Leu Thr Ser Leu						
			275		280			285
		Asn Gly Ala Leu Asp Pro Ile Met Tyr Phe Phe Val Ala Glu Lys Phe						
25			290		295			300
		Arg His Ala Leu Cys Asn Leu Leu Cys Gly Lys Arg Leu Lys Gly Pro						
			305		310		315	320
		Pro Pro Ser Phe Glu Gly Lys Thr Asn Glu Ser Ser Leu Ser Ala Lys						
			325		330			335
30		Ser Glu Leu						

(34) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

ATAAGATGAT CACCCTGAAC AATCAAGAT

29

(35) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

TCCGAATTCA TAACATTTCA CTGTTTATAT TGC

33

(36) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 996 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

20	ATGATCACCC TGAACAATCA AGATCAACCT GTCACCTTTTA ACAGCTCACA TCCAGATGAA	60
	TACAAAATTG CAGCCCTTGT CTTCTATAGC TGTATCTTCA TAATTGGATT ATTTGTTAAC	120
	ATCACTGCAT TATGGGTTTT CAGTTGTACC ACCAAGAAGA GAACCACGGT AACCATCTAT	180
	ATGATGAATG TGGCATTAGT GGA CTTGATA TTTATAATGA CTTTACCCTT TCGAATGTTT	240
	TATTATGCAA AAGATGCATG GCCATTTGGA GAGTACTTCT GCCAGATTAT TGGAGCTCTC	300
25	ACAGTGTTTT ACCCAAGCAT TGCTTTATGG CTTCTTGCCT TTATTAGTGC TGACAGATAC	360
	ATGGCCATTG TACAGCCGAA GTACGCCAAA GAACTTAAAA ACACGTGCAA AGCCGTGCTG	420
	GCGTGTGTGG GAGTCTGGAT AATGACCCTG ACCACGACCA CCCCTCTGCT ACTGCTCTAT	480
	AAAGACCCAG ATAAAGACTC CACTCCCGCC ACCTGCCTCA AGATTTCTGA CATCATCTAT	540
	CTAAAAGCTG TGAACGTGCT GAACCTCACT CGACTGACAT TTTTTTCTT GATTCCTTTG	600
30	TTCATCATGA TTGGGTGCTA CTTGGTCATT ATTCATAATC TCCTTCACGG CAGGACGTCT	660
	AAGCTGAAAC CCAAAGTCAA GGAGAAGTCC ATAAGGATCA TCATCACGCT GCTGGTGCAG	720

GTGCTCGTCT GCTTTATGCC CTTCACATC TGTTTCGCTT TCCTGATGCT GGGAACGGGG 780
 GAGAACAGTT ACAATCCCTG GGGAGCCTTT ACCACCTTCC TCATGAACCT CAGCACGTGT 840
 CTGGATGTGA TTCTCTACTA CATCGTTTCA AAACAATTTT AGGCTCGAGT CATTAGTGTC 900
 ATGCTATACC GTAATTACCT TCGAAGCCTG CGCAGAAAAA GTTTCCGATC TGGTAGTCTA 960
 5 AGGTCACTAA GCAATATAAA CAGTGAAATG TTATGA 996

(37) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 331 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

15 Met Ile Thr Leu Asn Asn Gln Asp Gln Pro Val Thr Phe Asn Ser Ser
 1 5 10 15
 His Pro Asp Glu Tyr Lys Ile Ala Ala Leu Val Phe Tyr Ser Cys Ile
 20 20 25 30
 Phe Ile Ile Gly Leu Phe Val Asn Ile Thr Ala Leu Trp Val Phe Ser
 35 40 45
 20 Cys Thr Thr Lys Lys Arg Thr Thr Val Thr Ile Tyr Met Met Asn Val
 50 55 60
 Ala Leu Val Asp Leu Ile Phe Ile Met Thr Leu Pro Phe Arg Met Phe
 65 70 75 80
 25 Tyr Tyr Ala Lys Asp Ala Trp Pro Phe Gly Glu Tyr Phe Cys Gln Ile
 85 90 95
 Ile Gly Ala Leu Thr Val Phe Tyr Pro Ser Ile Ala Leu Trp Leu Leu
 100 105 110
 Ala Phe Ile Ser Ala Asp Arg Tyr Met Ala Ile Val Gln Pro Lys Tyr
 115 120 125
 30 Ala Lys Glu Leu Lys Asn Thr Cys Lys Ala Val Leu Ala Cys Val Gly
 130 135 140
 Val Trp Ile Met Thr Leu Thr Thr Thr Thr Pro Leu Leu Leu Tyr
 145 150 155 160
 35 Lys Asp Pro Asp Lys Asp Ser Thr Pro Ala Thr Cys Leu Lys Ile Ser
 165 170 175

	Asp	Ile	Ile	Tyr	Leu	Lys	Ala	Val	Asn	Val	Leu	Asn	Leu	Thr	Arg	Leu	
					180					185					190		
	Thr	Phe	Phe	Phe	Leu	Ile	Pro	Leu	Phe	Ile	Met	Ile	Gly	Cys	Tyr	Leu	
					195					200					205		
5	Val	Ile	Ile	His	Asn	Leu	Leu	His	Gly	Arg	Thr	Ser	Lys	Leu	Lys	Pro	
					210					215					220		
	Lys	Val	Lys	Glu	Lys	Ser	Ile	Arg	Ile	Ile	Ile	Thr	Leu	Leu	Val	Gln	
					225					230					235		
	Val	Leu	Val	Cys	Phe	Met	Pro	Phe	His	Ile	Cys	Phe	Ala	Phe	Leu	Met	
10					245					250					255		
	Leu	Gly	Thr	Gly	Glu	Asn	Ser	Tyr	Asn	Pro	Trp	Gly	Ala	Phe	Thr	Thr	
					260					265					270		
	Phe	Leu	Met	Asn	Leu	Ser	Thr	Cys	Leu	Asp	Val	Ile	Leu	Tyr	Tyr	Ile	
					275					280					285		
15	Val	Ser	Lys	Gln	Phe	Gln	Ala	Arg	Val	Ile	Ser	Val	Met	Leu	Tyr	Arg	
					290					295					300		
	Asn	Tyr	Leu	Arg	Ser	Leu	Arg	Arg	Lys	Ser	Phe	Arg	Ser	Gly	Ser	Leu	
					305					310					315		
	Arg	Ser	Leu	Ser	Asn	Ile	Asn	Ser	Glu	Met	Leu						
20					325					330							

(38) INFORMATION FOR SEQ ID NO:37:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 28 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

CCAAGCTTCC AGGCCTGGGG TGTGCTGG

28

30 (39) INFORMATION FOR SEQ ID NO:38:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 29 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

35

ATGGATCCTG ACCTTCGGCC CCTGGCAGA

29

(40) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:

5

(A) LENGTH: 1077 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

10	ATGCCCTCTG TGTCTCCAGC GGGGCCCTCG GCCGGGGCAG TCCCCAATGC CACCGCAGTG	60
	ACAACAGTGC GGACCAATGC CAGCGGGCTG GAGGTGCCCC GTTCCACCT GTTTGCCCCG	120
	CTGGACGAGG AGCTGCATGG CACCTTCCCA GGCCTGTGCG TGGCGCTGAT GGCGGTGCAC	180
	GGAGCCATCT TCCTGGCAGG GCTGGTGCTC AACGGGCTGG CGCTGTACGT CTTCTGCTGC	240
	CGCACCCGGG CCAAGACACC CTCAGTCATC TACACCATCA ACCTGGTGGT GACCGATCTA	300
15	CTGGTAGGGC TGTCCCTGCC CACGCGCTTC GCTGTGTACT ACGGCGCCAG GGGCTGCCTG	360
	CGCTGTGCCT TCCCGCACGT CCTCGGTTAC TTCCTCAACA TGCACTGCTC CATCCTCTTC	420
	CTCACCTGCA TCTGCGTGGA CCGCTACCTG GCCATCGTGC GGCCCGAAGG CTCCCGCCGC	480
	TGCCGCCAGC CTGCCTGTGC CAGGGCCGTG TGCGCCTTCG TGTGGCTGGC CGCCGGTGCC	540
	GTCACCTGT CGGTGCTGGG CGTGACAGG AGCCGGCCCT GCTGCCGTGT CTTTGCGCTG	600
20	ACTGTCTTGG AGTTCCTGCT GCCCCTGCTG GTCATCAGCG TGTTTACCGG CCGCATCATG	660
	TGTGCACTGT CGCGGCCGGG TCTGCTCCAC CAGGGTCGCC AGCGCCGCGT GCGGGCCATG	720
	CAGCTCCTGC TCACGGTGCT CATCATCTTT CTCGTCTGCT TCACGCCCTT CCACGCCCGC	780
	CAAGTGGCCG TGGCGCTGTG GCCCACATG CCACACCACA CGAGCCTCGT GGTCTACCAC	840
	TGGGCCGTGA CCCTCAGCAG CCTCAACAGC TGCATGGACC CCATCGTCTA CTGCTTCGTC	900
25	ACCAGTGGCT TCCAGGCCAC CGTCCGAGGC CTCTTCGGCC AGCACGGAGA GCGTGAGCCC	960
	AGCAGCGGTG ACGTGGTCAG CATGCACAGG AGCTCCAAGG GCTCAGGCCG TCATCACATC	1020
	CTCAGTGCCG GCCCTCACGC CCTCACCAG GCCCTGGCTA ATGGGCCCCGA GGCTTAG	1077

(41) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

30

(A) LENGTH: 358 amino acids

(B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

	Met	Pro	Ser	Val	Ser	Pro	Ala	Gly	Pro	Ser	Ala	Gly	Ala	Val	Pro	Asn	
	1				5				10					15			
	Ala	Thr	Ala	Val	Thr	Thr	Val	Arg	Thr	Asn	Ala	Ser	Gly	Leu	Glu	Val	
				20				25						30			
10	Pro	Leu	Phe	His	Leu	Phe	Ala	Arg	Leu	Asp	Glu	Glu	Leu	His	Gly	Thr	
		35					40						45				
	Phe	Pro	Gly	Leu	Cys	Val	Ala	Leu	Met	Ala	Val	His	Gly	Ala	Ile	Phe	
		50					55					60					
15	Leu	Ala	Gly	Leu	Val	Leu	Asn	Gly	Leu	Ala	Leu	Tyr	Val	Phe	Cys	Cys	
	65				70				75						80		
	Arg	Thr	Arg	Ala	Lys	Thr	Pro	Ser	Val	Ile	Tyr	Thr	Ile	Asn	Leu	Val	
				85					90					95			
	Val	Thr	Asp	Leu	Leu	Val	Gly	Leu	Ser	Leu	Pro	Thr	Arg	Phe	Ala	Val	
				100				105						110			
20	Tyr	Tyr	Gly	Ala	Arg	Gly	Cys	Leu	Arg	Cys	Ala	Phe	Pro	His	Val	Leu	
		115					120					125					
	Gly	Tyr	Phe	Leu	Asn	Met	His	Cys	Ser	Ile	Leu	Phe	Leu	Thr	Cys	Ile	
		130				135						140					
25	Cys	Val	Asp	Arg	Tyr	Leu	Ala	Ile	Val	Arg	Pro	Glu	Ala	Pro	Ala	Ala	
	145				150					155				160			
	Cys	Arg	Gln	Pro	Ala	Cys	Ala	Arg	Ala	Val	Cys	Ala	Phe	Val	Trp	Leu	
				165				170						175			
	Ala	Ala	Gly	Ala	Val	Thr	Leu	Ser	Val	Leu	Gly	Val	Thr	Gly	Ser	Arg	
				180				185						190			
30	Pro	Cys	Cys	Arg	Val	Phe	Ala	Leu	Thr	Val	Leu	Glu	Phe	Leu	Leu	Pro	
		195					200					205					
	Leu	Leu	Val	Ile	Ser	Val	Phe	Thr	Gly	Arg	Ile	Met	Cys	Ala	Leu	Ser	
		210				215					220						
35	Arg	Pro	Gly	Leu	Leu	His	Gln	Gly	Arg	Gln	Arg	Arg	Val	Arg	Ala	Met	
	225				230					235				240			
	Gln	Leu	Leu	Leu	Thr	Val	Leu	Ile	Ile	Phe	Leu	Val	Cys	Phe	Thr	Pro	
				245				250					255				

Phe His Ala Arg Gln Val Ala Val Ala Leu Trp Pro Asp Met Pro His
260 265 270

His Thr Ser Leu Val Val Tyr His Val Ala Val Thr Leu Ser Ser Leu
275 280 285

S Asn Ser Cys Met Asp Pro Ile Val Tyr Cys Phe Val Thr Ser Gly Phe
290 295 300

Gln Ala Thr Val Arg Gly Leu Phe Gly Gln His Gly Glu Arg Glu Pro
305 310 315 320

[illegible]

Arg His His Ile Leu Ser Ala Gly Pro His Ala Leu Thr Gln Ala Leu
340 345 350

Ala Asn Gly Pro Glu Ala
355

15 (42) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GAGAATTCAC TCCTGAGCTC AAGATGAACT 30

(43) INFORMATION FOR SEO ID NO:42:

25 (i) SEQUENCE CHARACTERISTICS:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

30 (ii) MOLECULE TYPE: DNA (genomic)

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

CGGGATCCCC GTAAGTGAGC CACTTCAGAT 30

(44) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1050 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

ATGAACTCCA CCTGGATGG TAATCAGAGC AGCCACCCTT TTTGCCTCTT GGCATTGCGC 60
 5 TATTTGGAAA CTGTCAATTT TTGCCTTTTG GAAGTATTGA TTATTGTCTT TCTAACTGTA 120
 TTGATTATTT CTGGCAACAT CATTGTGATT TTTGTATTTC ACTGTGCACC TTTGTTGAAC 180
 CATCACACTA CAAGTTATTT TATCCAGACT ATGGCATATG CTGACCCTTT TGTGGGGTGG 240
 AGCTGCGTGG TCCCTTCTTT ATCACTCCTC CATCACCCCC TTCCAGTAGA GGAGTCCTTG 300
 ACTTGCCAGA TATTTGGTTT TGTAGTATCA GTTCTGAAGA GCGTCTCCAT GGCTTCTCTG 360
 10 GCCTGTATCA GCATTGATAG ATACATTGCC ATTACTAAAC CTTTAACCTA TAATACTCTG 420
 GTTACACCCT GGAGACTACG CCTGTGTATT TTCCTGATTT GGCTATACTC GACCCTGGTC 480
 TTCCTGCCTT CCTTTTCCA CTGGGGCAAA CCTGGATATC ATGGAGATGT GTTTCAGTGG 540
 TGTGCGGAGT CCTGGCACAC CGACTCCTAC TTCACCCTGT TCATCGTGAT GATGTTATAT 600
 GCCCCAGCAG CCCTTATTGT CTGCTTCACC TATTTCAACA TCTCCGCAT CTGCCAACAG 660
 15 CACACAAAGG ATATCAGCGA AAGGCAAGCC CGCTTCAGCA GCCAGAGTGG GGAGACTGGG 720
 GAAGTGCAGG CCTGTCCTGA TAAGCGCTAT GCCATGGTCC TGTTTCGAAT CACTAGTGTA 780
 TTTTACATCC TCTGGTTGCC ATATATCATC TACTTCTTGT TGGAAAGCTC CACTGGCCAC 840
 AGCAACCGCT TCGCATCCTT CTTGACCACC TGGCTTGCTA TTAGTAACAG TTTCTGCAAC 900
 TGTGTAATTT ATAGTCTCTC CAACAGTGTA TTCCAAAGAG GACTAAAGCG CCTCTCAGGG 960
 20 GCTATGTGTA CTTCTTGTGC AAGTCAGACT ACAGCCAACG ACCCTTACAC AGTTAGAAGC 1020
 AAAGGCCCTC TTAATGGATG TCATATCTGA 1050

(45) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:

25 (A) LENGTH: 349 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

30 Met Asn Ser Thr Leu Asp Gly Asn Gln Ser Ser His Pro Phe Cys Leu

35

	1		5		10		15									
	Leu	Ala	Phe	Gly	Tyr	Leu	Glu	Thr	Val	Asn	Phe	Cys	Leu	Leu	Glu	Val
				20					25						30	
5	Leu	Ile	Ile	Val	Phe	Leu	Thr	Val	Leu	Ile	Ile	Ser	Gly	Asn	Ile	Ile
			35					40					45			
	Val	Ile	Phe	Val	Phe	His	Cys	Ala	Pro	Leu	Leu	Asn	His	His	Thr	Thr
		50					55					60				
	Ser	Tyr	Phe	Ile	Gln	Thr	Met	Ala	Tyr	Ala	Asp	Leu	Phe	Val	Gly	Val
	65					70					75					80
10	Ser	Cys	Val	Val	Pro	Ser	Leu	Ser	Leu	Leu	His	His	Pro	Leu	Pro	Val
					85					90					95	
	Glu	Glu	Ser	Leu	Thr	Cys	Gln	Ile	Phe	Gly	Phe	Val	Val	Ser	Val	Leu
				100						105				110		
15	Lys	Ser	Val	Ser	Met	Ala	Ser	Leu	Ala	Cys	Ile	Ser	Ile	Asp	Arg	Tyr
			115					120					125			
	Ile	Ala	Ile	Thr	Lys	Pro	Leu	Thr	Tyr	Asn	Thr	Leu	Val	Thr	Pro	Trp
		130					135					140				
	Arg	Leu	Arg	Leu	Cys	Ile	Phe	Leu	Ile	Trp	Leu	Tyr	Ser	Thr	Leu	Val
	145					150					155					160
20	Phe	Leu	Pro	Ser	Phe	Phe	His	Trp	Gly	Lys	Pro	Gly	Tyr	His	Gly	Asp
					165					170					175	
	Val	Phe	Gln	Trp	Cys	Ala	Glu	Ser	Trp	His	Thr	Asp	Ser	Tyr	Phe	Thr
				180					185					190		
25	Leu	Phe	Ile	Val	Met	Met	Leu	Tyr	Ala	Pro	Ala	Ala	Leu	Ile	Val	Cys
				195				200					205			
	Phe	Thr	Tyr	Phe	Asn	Ile	Phe	Arg	Ile	Cys	Gln	Gln	His	Thr	Lys	Asp
		210					215					220				
	Ile	Ser	Glu	Arg	Gln	Ala	Arg	Phe	Ser	Ser	Gln	Ser	Gly	Glu	Thr	Gly
	225					230					235					240
30	Glu	Val	Gln	Ala	Cys	Pro	Asp	Lys	Arg	Tyr	Ala	Met	Val	Leu	Phe	Arg
				245						250					255	
	Ile	Thr	Ser	Val	Phe	Tyr	Ile	Leu	Trp	Leu	Pro	Tyr	Ile	Ile	Tyr	Phe
				260					265					270		
35	Leu	Leu	Glu	Ser	Ser	Thr	Gly	His	Ser	Asn	Arg	Phe	Ala	Ser	Phe	Leu
			275					280					285			
	Thr	Thr	Trp	Leu	Ala	Ile	Ser	Asn	Ser	Phe	Cys	Asn	Cys	Val	Ile	Tyr
		290					295					300				

Ser Leu Ser Asn Ser Val Phe Gln Arg Gly Leu Lys Arg Leu Ser Gly
305 310 315 320

Ala Met Cys Thr Ser Cys Ala Ser Gln Thr Thr Ala Asn Asp Pro Tyr
325 330 335

5 Thr Val Arg Ser Lys Gly Pro Leu Asn Gly Cys His Ile
340 345

(46) INFORMATION FOR SEQ ID NO:45:

- 10 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

15 TCCCCCGGGA AAAAAACCAA CTGCTCCAAA 30

(47) INFORMATION FOR SEQ ID NO:46:

- 20 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

TAGGATCCAT TTGAATGTGG ATTTGGTGAA A 31

25 (48) INFORMATION FOR SEQ ID NO:47:

- 30 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1302 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

ATGTGTTTTT CTCCCATTCT GGAAATCAAC ATGCAGTCTG AATCTAACAT TACAGTGCGA 60

GATGACATTG ATGACATCAA CACCAATATG TACCAACCAC TATCATATCC GTTAAGCTTT 120

35 CAAGTGTCTC TCACCGGATT TCTTATGTTA GAAATTGTGT TGGGACTTGG CAGCAACCTC 180

	ACTGTATTGG	TACTTTACTG	CATGAAATCC	AACTTAATCA	ACTCTGTCAG	TAACATTATT	240
	ACAATGAATC	TTCATGTACT	TGATGTAATA	ATTTGTGTGG	GATGTATTCC	TCTAACTATA	300
	GTTATCCTTC	TGCTTTCACT	GGAGAGTAAC	ACTGCTCTCA	TTTGCTGTTT	CCATGAGGCT	360
	TGTGTATCTT	TTGCAAGTGT	CTCAACAGCA	ATCAACGTTT	TTGCTATCAC	TTTGGACAGA	420
5	TATGACATCT	CTGTAAACC	TGCAAACCGA	ATTCTGACAA	TGGGCAGAGC	TGTAATGTTA	480
	ATGATATCCA	TTTGGATTTT	TTCTTTTTTC	TCTTTCCTGA	TTCTTTTAT	TGAGGTAAAT	540
	TTTTTCAGTC	TTCAAAGTGG	AAATACCTGG	GAAAACAAGA	CACTTTTATG	TGTCAGTACA	600
	AATGAATACT	ACACTGAACT	GGGAATGTAT	TATCACCTGT	TAGTACAGAT	CCCAATATTC	660
	TTTTTCACTG	TTGTAGTAAT	GTTAATCACA	TACACCAAAA	TACTTCAGGC	TCTTAATATT	720
10	CGAATAGGCA	CAAGATTTTC	AACAGGGCAG	AAGAAGAAAG	CAAGAAAGAA	AAAGACAATT	780
	TCTCTAACCA	CACAACATGA	GGCTACAGAC	ATGTCACAAA	GCAGTGGTGG	GAGAAATGTA	840
	GTCTTTGGTG	TAAGAACTTC	AGTTTCTGTA	ATAATTGCCC	TCCGGCGAGC	TGTGAAACGA	900
	CACCGTGAAC	GACGAGAAAG	ACAAAAGAGA	GTCTTCAGGA	TGTCTTTATT	GATTATTTCT	960
	ACATTTCTTC	TCTGCTGGAC	ACCAATTTCT	GTTTTAAATA	CCACCATTTT	ATGTTTAGGC	1020
15	CCAAGTGACC	TTTTAGTAAA	ATTAAGATTG	TGTTTTTTTAG	TCATGGCTTA	TGGAACAACCT	1080
	ATATTTCAAC	CTCTATTATA	TGCATTCACT	AGACAAAAAT	TTCAAAAGGT	CTTGAAAAGT	1140
	AAAATGAAAA	AGCGAGTTGT	TTCTATAGTA	GAAGCTGATC	CCCTGCCTAA	TAATGCTGTA	1200
	ATACACAACT	CTTGGATAGA	TCCCAAAAGA	AACAAAAAAA	TTACCTTTGA	AGATAGTGAA	1260
	ATAAGAGAAA	AACGTTTAGT	GCCTCAGGTT	GTCACAGACT	AG		1302

20 (49) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 433 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

Met Cys Phe Ser Pro Ile Leu Glu Ile Asn Met Gln Ser Glu Ser Asn
1 5 10 15

30 Ile Thr Val Arg Asp Asp Ile Asp Asp Ile Asn Thr Asn Met Tyr Gln
 20 25 30

QMS777IN -WD 002212041 1 -

[illegible]

5 Leu Val Met Ala Tyr Gly Thr Thr Ile Phe His Pro Leu Leu Tyr Ala
 355 360 365

Arg Val Val Ser Ile Val Glu Ala Asp Pro Leu Pro Asn Asn Ala Val
385 390 395 400

Glu Asp Ser Glu Ile Arg Glu Lys Arg Leu Val Pro Gln Val Val Thr
420 425 430

15

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

30

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

31

(A) LENGTH: 1209 base pairs

(B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

	ATGTTGTGTC CTTCCAAGAC AGATGGCTCA GGGCACTCTG GTAGGATTCA CCAGGAAACT	60
	CATGGAGAAG GGAAAAGGGA CAAGATTAGC AACAGTGAAG GGAGGGAGAA TGGTGGGAGA	120
	GGATTCCAGA TGAACGGTGG GTCGCTGGAG GCTGAGCATG CCAGCAGGAT GTCAGTTCTC	180
	AGAGCAAAGC CCATGTCAAA CAGCCAACGC TTGCTCCTTC TGTCCCCAGG ATCACCTCCT	240
10	CGCACGGGGA GCATCTCCTA CATCAACATC ATCATGCCTT CGGTGTTCGG CACCATCTGC	300
	CTCCTGGGCA TCATCGGGAA CTCCACGGTC ATCTTCGCGG TCGTGAAGAA GTCCAAGCTG	360
	CACTGGTGCA ACAACGTCCC CGACATCTTC ATCATCAACC TCTCGGTAGT AGATCTCCTC	420
	TTTCTCCTGG GCATGCCCTT CATGATCCAC CAGCTCATGG GCAATGGGGT GTGGCACTTT	480
	GGGGAGACCA TGTGCACCTT CATCACGGCC ATGGATGCCA ATAGTCAGTT CACCAGCACC	540
15	TACATCCTGA CCGCCATGGC CATTGACCGC TACCTGGCCA CTGTCCACCC CATCTCTTCC	600
	ACGAAGTTCC GGAAGCCCTC TGTGGCCACC CTGGTGATCT GCCTCCTGTG GGCCCTCTCC	660
	TTCATCAGCA TCACCCCTGT GTGGCTGTAT GCCAGACTCA TCCCCTTCCC AGGAGGTGCA	720
	GTGGGCTGCG GCATACGCCT GCCCAACCCA GACACTGACC TCTACTGGTT CACCCTGTAC	780
	CAGTTTTTCC TGGCCTTTGC CCTGCCTTTT GTGGTCATCA CAGCCGCATA CGTGAGGATC	840
20	CTGCAGCGCA TGACGTCTTC AGTGGCCCCC GCCTCCCAGC GCAGCATCCG GCTGCGGACA	900
	AAGAGGTGA CCCGCACAGC CATCGCCATC TGTCTGGTCT TCTTTGTGTG CTGGGCACCC	960
	TACTATGTGC TACAGCTGAC CCAGTTGTCC ATCAGCCGCC CGACCCTCAC CTTTGTCTAC	1020
	TTATACAATG CGGCCATCAG CTTGGGCTAT GCCAACAGCT GCCTCAACCC CTTTGTGTAC	1080
	ATCGTGCTCT GTGAGACGTT CCGCAAACGC TTGGTCCTGT CGGTGAAGCC TGCAGCCCAG	1140
25	GGGCAGCTTC GCGCTGTCAG CAACGCTCAG ACGGCTGACG AGGAGAGGAC AGAAAGCAAA	1200
	GGCACCTGA	1209

(53) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 402 amino acids
 (B) TYPE: amino acid

30

(C) STRANDEDNESS:

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

5 Met Leu Cys Pro Ser Lys Thr Asp Gly Ser Gly His Ser Gly Arg Ile
 1 5 10 15
 His Gln Glu Thr His Gly Glu Gly Lys Arg Asp Lys Ile Ser Asn Ser
 20 25 30
 10 Glu Gly Arg Glu Asn Gly Gly Arg Gly Phe Gln Met Asn Gly Gly Ser
 35 40 45
 Leu Glu Ala Glu His Ala Ser Arg Met Ser Val Leu Arg Ala Lys Pro
 50 55 60
 Met Ser Asn Ser Gln Arg Leu Leu Leu Leu Ser Pro Gly Ser Pro Pro
 65 70 75 80
 15 Arg Thr Gly Ser Ile Ser Tyr Ile Asn Ile Ile Met Pro Ser Val Phe
 85 90 95
 Gly Thr Ile Cys Leu Leu Gly Ile Ile Gly Asn Ser Thr Val Ile Phe
 100 105 110
 20 Ala Val Val Lys Lys Ser Lys Leu His Trp Cys Asn Asn Val Pro Asp
 115 120 125
 Ile Phe Ile Ile Asn Leu Ser Val Val Asp Leu Leu Phe Leu Leu Gly
 130 135 140
 Met Pro Phe Met Ile His Gln Leu Met Gly Asn Gly Val Trp His Phe
 145 150 155 160
 25 Gly Glu Thr Met Cys Thr Leu Ile Thr Ala Met Asp Ala Asn Ser Gln
 165 170 175
 Phe Thr Ser Thr Tyr Ile Leu Thr Ala Met Ala Ile Asp Arg Tyr Leu
 180 185 190
 30 Ala Thr Val His Pro Ile Ser Ser Thr Lys Phe Arg Lys Pro Ser Val
 195 200 205
 Ala Thr Leu Val Ile Cys Leu Leu Trp Ala Leu Ser Phe Ile Ser Ile
 210 215 220
 Thr Pro Val Trp Leu Tyr Ala Arg Leu Ile Pro Phe Pro Gly Gly Ala
 225 230 235 240
 35 Val Gly Cys Gly Ile Arg Leu Pro Asn Pro Asp Thr Asp Leu Tyr Trp
 245 250 255

Phe Thr Leu Tyr Gln Phe Phe Leu Ala Phe Ala Leu Pro Phe Val Val
 260 265 270
 Ile Thr Ala Ala Tyr Val Arg Ile Leu Gln Arg Met Thr Ser Ser Val
 275 280 285
 5 Ala Pro Ala Ser Gln Arg Ser Ile Arg Leu Arg Thr Lys Arg Val Thr
 290 295 300
 Arg Thr Ala Ile Ala Ile Cys Leu Val Phe Phe Val Cys Trp Ala Pro
 305 310 315 320
 10 Tyr Tyr Val Leu Gln Leu Thr Gln Leu Ser Ile Ser Arg Pro Thr Leu
 325 330 335
 Thr Phe Val Tyr Leu Tyr Asn Ala Ala Ile Ser Leu Gly Tyr Ala Asn
 340 345 350
 Ser Cys Leu Asn Pro Phe Val Tyr Ile Val Leu Cys Glu Thr Phe Arg
 355 360 365
 15 Lys Arg Leu Val Leu Ser Val Lys Pro Ala Ala Gln Gly Gln Leu Arg
 370 375 380
 Ala Val Ser Asn Ala Gln Thr Ala Asp Glu Glu Arg Thr Glu Ser Lys
 385 390 395 400
 Gly Thr

20

(54) INFORMATION FOR SEQ ID NO:53:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 27 base pairs
 (B) TYPE: nucleic acid
 25 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

 (ii) MOLECULE TYPE: DNA (genomic)

 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

GGCGGATCCA TGGATGTGAC TTCCCAA

27

30

(55) INFORMATION FOR SEQ ID NO:54:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 27 base pairs
 (B) TYPE: nucleic acid
 35 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

 (ii) MOLECULE TYPE: DNA (genomic)

 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

GGCGGATCCC TACACGGCAC TGCTGAA

27

(56) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:

5

- (A) LENGTH: 1128 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

10	ATGGATGTGA CTTCCAAGC CCGGGGCGTG GGCCTGGAGA TGTACCCAGG CACCGCGCAC	60
	GCTGCGGCCC CCAACACCAC CTCCCCCGAG CTCAACCTGT CCCACCCGCT CCTGGGCACC	120
	GCCCTGGCCA ATGGGACAGG TGAGCTCTCG GAGCACCAGC AGTACGTGAT CGGCCTGTTC	180
	CTCTCGTGCC TCTACACCAT CTTCTCTTC CCCATCGGCT TTGTGGGCAA CATCCTGATC	240
	CTGGTGGTGA ACATCAGCTT CCGCGAGAAG ATGACCATCC CCGACCTGTA CTTTCATCAAC	300
15	CTGGCGGTGG CGGACCTCAT CCTGGTGGCC GACTCCCTCA TTGAGGTGTT CAACCTGCAC	360
	GAGCGGTACT ACGACATCGC CGTCCTGTGC ACCTTCATGT CGCTCTTCCT GCAGGTCAAC	420
	ATGTACAGCA GCGTCTTCTT CCTCACCTGG ATGAGCTTCG ACCGCTACAT CGCCCTGGCC	480
	AGGGCCATGC GCTGCAGCCT GTTCCGCACC AAGCACCACG CCCGGCTGAG CTGTGGCCTC	540
	ATCTGGATGG CATCCGTGTC AGCCACGCTG GTGCCCTTCA CCGCCGTGCA CCTGCAGCAC	600
20	ACCGACGAGG CTGCTTCTG TTTTCGCGGAT GTCCGGGAGG TGCAGTGGCT CGAGGTCACG	660
	CTGGGCTTCA TCGTGCCCTT CGCCATCATC GGCCTGTGCT ACTCCCTCAT TGTCCGGGTG	720
	CTGGTCAGGG CGCACCGGCA CCGTGGGCTG CGGCCCCGGC GGCAGAAGGC GCTCCGCATG	780
	ATCCTCGCGG TGGTGCTGGT CTTCTTCGTC TGCTGGCTGC CGGAGAACGT CTTTCATCAGC	840
	GTGCACCTCC TGCAGCGGAC GCAGCCTGGG GCCGCTCCCT GCAAGCAGTC TTTCCGCCAT	900
25	GCCCACCCCC TCACGGGCCA CATTGTCAAC CTCGCCGCCT TCTCCAACAG CTGCCTAAAC	960
	CCCCTCATCT ACAGCTTTCT CGGGGAGACC TTCAGGGACA AGCTGAGGCT GTACATTGAG	1020
	CAGAAACAA ATTTGCCGGC CCTGAACCGC TTCTGTCACG CTGCCCTGAA GGCCGTCATT	1080
	CCAGACAGCA CCGAGCAGTC GGATGTGAGG TTCAGCAGTG CCGTGTGA	1128

(57) INFORMATION FOR SEQ ID NO:56:

30

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 375 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

5 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

	Met	Asp	Val	Thr	Ser	Gln	Ala	Arg	Gly	Val	Gly	Leu	Glu	Met	Tyr	Pro	
	1				5					10					15		
10	Gly	Thr	Ala	His	Ala	Ala	Ala	Pro	Asn	Thr	Thr	Ser	Pro	Glu	Leu	Asn	
				20					25					30			
	Leu	Ser	His	Pro	Leu	Leu	Gly	Thr	Ala	Leu	Ala	Asn	Gly	Thr	Gly	Glu	
			35					40					45				
	Leu	Ser	Glu	His	Gln	Gln	Tyr	Val	Ile	Gly	Leu	Phe	Leu	Ser	Cys	Leu	
		50					55					60					
15	Tyr	Thr	Ile	Phe	Leu	Phe	Pro	Ile	Gly	Phe	Val	Gly	Asn	Ile	Leu	Ile	
	65					70					75				80		
	Leu	Val	Val	Asn	Ile	Ser	Phe	Arg	Glu	Lys	Met	Thr	Ile	Pro	Asp	Leu	
				85						90					95		
20	Tyr	Phe	Ile	Asn	Leu	Ala	Val	Ala	Asp	Leu	Ile	Leu	Val	Ala	Asp	Ser	
				100					105					110			
	Leu	Ile	Glu	Val	Phe	Asn	Leu	His	Glu	Arg	Tyr	Tyr	Asp	Ile	Ala	Val	
			115					120					125				
	Leu	Cys	Thr	Phe	Met	Ser	Leu	Phe	Leu	Gln	Val	Asn	Met	Tyr	Ser	Ser	
		130					135					140					
25	Val	Phe	Phe	Leu	Thr	Trp	Met	Ser	Phe	Asp	Arg	Tyr	Ile	Ala	Leu	Ala	
	145					150					155				160		
	Arg	Ala	Met	Arg	Cys	Ser	Leu	Phe	Arg	Thr	Lys	His	His	Ala	Arg	Leu	
				165						170					175		
30	Ser	Cys	Gly	Leu	Ile	Trp	Met	Ala	Ser	Val	Ser	Ala	Thr	Leu	Val	Pro	
				180					185					190			
	Phe	Thr	Ala	Val	His	Leu	Gln	His	Thr	Asp	Glu	Ala	Cys	Phe	Cys	Phe	
			195				200						205				
	Ala	Asp	Val	Arg	Glu	Val	Gln	Trp	Leu	Glu	Val	Thr	Leu	Gly	Phe	Ile	
		210					215					220					
35	Val	Pro	Phe	Ala	Ile	Ile	Gly	Leu	Cys	Tyr	Ser	Leu	Ile	Val	Arg	Val	
	225					230					235				240		
	Leu	Val	Arg	Ala	His	Arg	His	Arg	Gly	Leu	Arg	Pro	Arg	Arg	Gln	Lys	

45

		245		250		255										
	Ala	Leu	Arg	Met	Ile	Leu	Ala	Val	Val	Leu	Val	Phe	Phe	Val	Cys	Trp
		260						265						270		
5	Leu	Pro	Glu	Asn	Val	Phe	Ile	Ser	Val	His	Leu	Leu	Gln	Arg	Thr	Gln
		275						280					285			
	Pro	Gly	Ala	Ala	Pro	Cys	Lys	Gln	Ser	Phe	Arg	His	Ala	His	Pro	Leu
		290					295					300				
	Thr	Gly	His	Ile	Val	Asn	Leu	Ala	Ala	Phe	Ser	Asn	Ser	Cys	Leu	Asn
	305					310					315				320	
10	Pro	Leu	Ile	Tyr	Ser	Phe	Leu	Gly	Glu	Thr	Phe	Arg	Asp	Lys	Leu	Arg
					325					330					335	
	Leu	Tyr	Ile	Glu	Gln	Lys	Thr	Asn	Leu	Pro	Ala	Leu	Asn	Arg	Phe	Cys
				340				345					350			
15	His	Ala	Ala	Leu	Lys	Ala	Val	Ile	Pro	Asp	Ser	Thr	Glu	Gln	Ser	Asp
		355					360						365			
	Val	Arg	Phe	Ser	Ser	Ala	Val									
		370					375									

(58) INFORMATION FOR SEQ ID NO:57:

- 20 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 31 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

AAGGAATTCA CGGCCGGGTG ATGCCATTCC C

31

(59) INFORMATION FOR SEQ ID NO:58:

- 30 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

35 GGTGGATCCA TAAACACGGG CGTTGAGGAC

30

(60) INFORMATION FOR SEQ ID NO:59:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 960 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

5

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

```

ATGCCATTCC CAAACTGCTC AGCCCCCAGC ACTGTGGTGG CCACAGCTGT GGGTGTCTTG      60
CTGGGGCTGG AGTGTGGGCT GGGTCTGCTG GGCAACGCGG TGGCGCTGTG GACCTTCCTG      120
10 TTCCGGGTCA GGGTGTGGAA GCCGTACGCT GTCTACCTGC TCAACCTGGC CCTGGCTGAC      180
CTGCTGTTGG CTGCGTGCCT GCCTTTCCTG GCCGCCTTCT ACCTGAGCCT CCAGGCTTGG      240
CATCTGGGCC GTGTGGGCTG CTGGGCCCTG CGCTTCCTGC TGGACCTCAG CCGCAGCGTG      300
GGGATGGCCT TCCTGGCCGC CGTGGCTTTG GACCGGTACC TCCGTGTGGT CCACCCTCGG      360
CTTAAGGTCA ACCTGCTGTC TCCTCAGGCG GCCCTGGGGG TCTCGGGCCT CGTCTGGCTC      420
15 CTGATGGTCG CCCTCACCTG CCCGGGCTTG CTCATCTCTG AGGCCGCCCA GAACTCCACC      480
AGGTGCCACA GTTTCTACTC CAGGGCAGAC GGCTCCTTCA GCATCATCTG GCAGGAAGCA      540
CTCTCCTGCC TTCAGTTTGT CCTCCCCCTT GGCCTCATCG TGTCTGCAA TGCAGGCATC      600
ATCAGGGCTC TCCAGAAAAG ACTCCGGGAG CTTGAGAAAC AGCCCAAGCT TCAGCGGGCC      660
CAGGCACTGG TCACCTTGGT GGTGGTGCTG TTTGCTCTGT GCTTTCTGCC CTGCTTCCTG      720
20 GCCAGAGTCC TGATGCACAT CTTCCAGAAT CTGGGGAGCT GCAGGGCCCT TTGTGCAGTG      780
GCTCATACCT CGGATGTCAC GGGCAGCCTC ACCTACCTGC ACAGTGTCTG CAACCCCGTG      840
GTATACTGCT TCTCCAGCCC CACCTTCAGG AGCTCCTATC GGAGGGTCTT CCACACCCTC      900
CGAGGCAAAG GGCAGGCAGC AGAGCCCCCA GATTTCACC CCAGAGACTC CTATTCCTGA      960

```

(61) INFORMATION FOR SEQ ID NO:60:

- 25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 319 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

30 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

Met Pro Phe Pro Asn Cys Ser Ala Pro Ser Thr Val Val Ala Thr Ala

47

	1		5		10		15									
	Val	Gly	Val	Leu	Leu	Gly	Leu	Glu	Cys	Gly	Leu	Gly	Leu	Leu	Gly	Asn
				20				25						30		
5	Ala	Val	Ala	Leu	Trp	Thr	Phe	Leu	Phe	Arg	Val	Arg	Val	Trp	Lys	Pro
			35					40					45			
	Tyr	Ala	Val	Tyr	Leu	Leu	Asn	Leu	Ala	Leu	Ala	Asp	Leu	Leu	Leu	Ala
		50					55					60				
	Ala	Cys	Leu	Pro	Phe	Leu	Ala	Ala	Phe	Tyr	Leu	Ser	Leu	Gln	Ala	Trp
	65					70					75				80	
10	His	Leu	Gly	Arg	Val	Gly	Cys	Trp	Ala	Leu	Arg	Phe	Leu	Leu	Asp	Leu
					85					90					95	
	Ser	Arg	Ser	Val	Gly	Met	Ala	Phe	Leu	Ala	Ala	Val	Ala	Leu	Asp	Arg
				100					105					110		
15	Tyr	Leu	Arg	Val	Val	His	Pro	Arg	Leu	Lys	Val	Asn	Leu	Leu	Ser	Pro
			115					120					125			
	Gln	Ala	Ala	Leu	Gly	Val	Ser	Gly	Leu	Val	Trp	Leu	Leu	Met	Val	Ala
		130					135					140				
	Leu	Thr	Cys	Pro	Gly	Leu	Leu	Ile	Ser	Glu	Ala	Ala	Gln	Asn	Ser	Thr
	145					150					155					160
20	Arg	Cys	His	Ser	Phe	Tyr	Ser	Arg	Ala	Asp	Gly	Ser	Phe	Ser	Ile	Ile
					165					170					175	
	Trp	Gln	Glu	Ala	Leu	Ser	Cys	Leu	Gln	Phe	Val	Leu	Pro	Phe	Gly	Leu
				180					185					190		
25	Ile	Val	Phe	Cys	Asn	Ala	Gly	Ile	Ile	Arg	Ala	Leu	Gln	Lys	Arg	Leu
			195					200					205			
	Arg	Glu	Pro	Glu	Lys	Gln	Pro	Lys	Leu	Gln	Arg	Ala	Gln	Ala	Leu	Val
		210					215					220				
	Thr	Leu	Val	Val	Val	Leu	Phe	Ala	Leu	Cys	Phe	Leu	Pro	Cys	Phe	Leu
	225					230					235					240
30	Ala	Arg	Val	Leu	Met	His	Ile	Phe	Gln	Asn	Leu	Gly	Ser	Cys	Arg	Ala
					245					250					255	
	Leu	Cys	Ala	Val	Ala	His	Thr	Ser	Asp	Val	Thr	Gly	Ser	Leu	Thr	Tyr
				260					265					270		
35	Leu	His	Ser	Val	Val	Asn	Pro	Val	Val	Tyr	Cys	Phe	Ser	Ser	Pro	Thr
			275					280					285			
	Phe	Arg	Ser	Ser	Tyr	Arg	Arg	Val	Phe	His	Thr	Leu	Arg	Gly	Lys	Gly
		290					295						300			

Gln Ala Ala Glu Pro Pro Asp Phe Asn Pro Arg Asp Ser Tyr Ser
 305 310 315

(62) INFORMATION FOR SEQ ID NO:61:

- 5 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1143 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

	ATGGAGGAAG GTGGTGATTT TGACAACTAC TATGGGGCAG ACAACCAGTC TGAGTGTGAG	60
	TACACAGACT GGAAATCCTC GGGGGCCCTC ATCCCTGCCA TCTACATGTT GGTCTTCCTC	120
	CTGGGCACCA CGGGAAACGG TCTGGTGCTC TGGACCGTGT TTCGGAGCAG CCGGGAGAAG	180
	AGGCGCTCAG CTGATATCTT CATTGCTAGC CTGGCGGTGG CTGACCTGAC CTTCGTGGTG	240
15	ACGCTGCCCC TGTGGGCTAC CTACACGTAC CGGGACTATG ACTGGCCCTT TGGGACCTTC	300
	TTCTGCAAGC TCAGCAGCTA CCTCATCTTC GTCAACATGT ACGCCAGCGT CTTCTGCCTC	360
	ACCGGCCTCA GCTTCGACCG CTACCTGGCC ATCGTGAGGC CAGTGGCCAA TGCTCGGCTG	420
	AGGCTGCGGG TCAGCGGGGC CGTGGCCACG GCAGTTCTTT GGGTGCTGGC CGCCCTCCTG	480
	GCCATGCCTG TCATGGTGTT ACGCACCACC GGGGACTTGG AGAACACCAC TAAGGTGCAG	540
20	TGCTACATGG ACTACTCCAT GGTGGCCACT GTGAGCTCAG AGTGGGCCTG GGAGGTGGGC	600
	CTTGGGGTCT CGTCCACCAC CGTGGGCTTT GTGGTGCCCT TCACCATCAT GCTGACCTGT	660
	TACTTCTTCA TCGCCCAAAC CATCGCTGGC CACTTCCGCA AGGAACGCAT CGAGGGCCTG	720
	CGGAAGCGGC GCCGGCTGCT CAGCATCATC GTGGTGCTGG TGGTGACCTT TGCCCTGTGC	780
	TGGATGCCCT ACCACCTGGT GAAGACGCTG TACATGCTGG GCAGCCTGCT GCACTGGCCC	840
25	TGTGACTTTG ACCTCTTCCT CATGAACATC TTCCCCTACT GCACCTGCAT CAGCTACGTC	900
	AACAGCTGCC TCAACCCCTT CCTCTATGCC TTTTTCGACC CCCGCTTCCG CCAGGCCTGC	960
	ACCTCCATGC TCTGCTGTGG CCAGAGCAGG TGCGCAGGCA CCTCCCACAG CAGCAGTGGG	1020
	GAGAAGTCAG CCAGCTACTC TTCGGGGCAC AGCCAGGGGC CCGGCCCAA CATGGGCAAG	1080
	GGTGGAGAAC AGATGCACGA GAAATCCATC CCCTACAGCC AGGAGACCCT TGTGGTTGAC	1140
30	TAG	1143

(63) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 380 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

5 Met Glu Glu Gly Gly Asp Phe Asp Asn Tyr Tyr Gly Ala Asp Asn Gln
 1 5 10 15
 Ser Glu Cys Glu Tyr Thr Asp Trp Lys Ser Ser Gly Ala Leu Ile Pro
 20 25 30
 Ala Ile Tyr Met Leu Val Phe Leu Leu Gly Thr Thr Gly Asn Gly Leu
 35 40 45
 15 Val Leu Trp Thr Val Phe Arg Ser Ser Arg Glu Lys Arg Arg Ser Ala
 50 55 60
 Asp Ile Phe Ile Ala Ser Leu Ala Val Ala Asp Leu Thr Phe Val Val
 65 70 75 80
 20 Thr Leu Pro Leu Trp Ala Thr Tyr Thr Tyr Arg Asp Tyr Asp Trp Pro
 85 90 95
 Phe Gly Thr Phe Phe Cys Lys Leu Ser Ser Tyr Leu Ile Phe Val Asn
 100 105 110
 Met Tyr Ala Ser Val Phe Cys Leu Thr Gly Leu Ser Phe Asp Arg Tyr
 115 120 125
 25 Leu Ala Ile Val Arg Pro Val Ala Asn Ala Arg Leu Arg Leu Arg Val
 130 135 140
 Ser Gly Ala Val Ala Thr Ala Val Leu Trp Val Leu Ala Ala Leu Leu
 145 150 155 160
 30 Ala Met Pro Val Met Val Leu Arg Thr Thr Gly Asp Leu Glu Asn Thr
 165 170 175
 Thr Lys Val Gln Cys Tyr Met Asp Tyr Ser Met Val Ala Thr Val Ser
 180 185 190
 Ser Glu Trp Ala Trp Glu Val Gly Leu Gly Val Ser Ser Thr Thr Val
 195 200 205
 35 Gly Phe Val Val Pro Phe Thr Ile Met Leu Thr Cys Tyr Phe Phe Ile
 210 215 220
 Ala Gln Thr Ile Ala Gly His Phe Arg Lys Glu Arg Ile Glu Gly Leu

	225		230		235		240
	Arg Lys Arg Arg Arg Leu Leu Ser Ile Ile Val Val Leu Val Val Thr	245		250		255	
5	Phe Ala Leu Cys Trp Met Pro Tyr His Leu Val Lys Thr Leu Tyr Met	260		265		270	
	Leu Gly Ser Leu Leu His Trp Pro Cys Asp Phe Asp Leu Phe Leu Met	275		280		285	
	Asn Ile Phe Pro Tyr Cys Thr Cys Ile Ser Tyr Val Asn Ser Cys Leu	290		295		300	
10	Asn Pro Phe Leu Tyr Ala Phe Phe Asp Pro Arg Phe Arg Gln Ala Cys	305		310		315	320
	Thr Ser Met Leu Cys Cys Gly Gln Ser Arg Cys Ala Gly Thr Ser His	325		330		335	
15	Ser Ser Ser Gly Glu Lys Ser Ala Ser Tyr Ser Ser Gly His Ser Gln	340		345		350	
	Gly Pro Gly Pro Asn Met Gly Lys Gly Gly Glu Gln Met His Glu Lys	355		360		365	
	Ser Ile Pro Tyr Ser Gln Glu Thr Leu Val Val Asp	370		375		380	

20 (64) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

TGAGAATTCT GGTGACTCAC AGCCGGCACA G

31

(65) INFORMATION FOR SEQ ID NO:64:

30 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

35 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

GCCGGATCCA AGGAAAAGCA GCAATAAAAG G

31

(66) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1119 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

10 ATGAACTACC CGCTAACGCT GGAAATGGAC CTCGAGAACC TGGAGGACCT GTTCTGGGAA 60
CTGGACAGAT TGGACAACATA TAACGACACC TCCCTGGTGG AAAATCATCT CTGCCCTGCC 120
ACAGAGGGTC CCCTCATGGC CTCCTTCAAG GCCGTGTTCG TGCCCGTGGC CTACAGCCTC 180
ATCTTCCTCC TGGGCGTGAT CGGCAACGTC CTGGTGCTGG TGATCCTGGA GCGGCACCGG 240
CAGACACGCA GTTCCACGGA GACCTTCCTG TTCCACCTGG CCGTGGCCGA CCTCCTGCTG 300
15 GTCTTCATCT TGCCCTTTGC CGTGGCCGAG GGCTCTGTGG GCTGGGTCCT GGGGACCTTC 360
CTCTGCAAAA CTGTGATTGC CCTGCACAAA GTCAACTTCT ACTGCAGCAG CCTGCTCCTG 420
GCCTGCATCG CCGTGGACCG CTACCTGGCC ATTGTCCACG CCGTCCATGC CTACCGCCAC 480
CGCCGCCTCC TCTCCATCCA CATCACCTGT GGGACCATCT GGCTGGTGGG CTTCTCTCCTT 540
GCCTTGCCAG AGATTCTCTT CGCCAAAGTC AGCCAAGGCC ATCACAACAA CTCCCTGCCA 600
20 CGTTGCACCT TCTCCAAGA GAACCAAGCA GAAACGCATG CCTGGTTCAC CTCCCGATTC 660
CTCTACCATG TGGCGGGATT CCTGCTGCCC ATGCTGGTGA TGGGCTGGTG CTACGTGGGG 720
GTAGTGACA GGTTCGCGCA GGCCAGCGG CGCCCTCAGC GGCAGAAGGC AGTCAGGGTG 780
GCCATCCTGG TGACAAGCAT CTTCTTCCTC TGCTGGTCAC CCTACCACAT CGTCATCTTC 840
CTGGACACCC TGGCGAGGCT GAAGGCCGTG GACAATACCT GCAAGCTGAA TGGCTCTCTC 900
25 CCCGTGGCCA TCACCATGTG TGAGTTCCTG GGCCTGGCCC ACTGCTGCCT CAACCCCATG 960
CTCTACACTT TCGCCGGCGT GAAGTTCGCG AGTGACCTGT CGCGGCTCCT GACCAAGCTG 1020
GGCTGTACCG GCCCTGCCTC CCTGTGCCAG CTCTTCCCTA GCTGGCGCAG GAGCAGTCTC 1080
TCTGAGTCAG AGAATGCCAC CTCTCTCACC ACGTTCTAG 1119

(67) INFORMATION FOR SEQ ID NO:66:

30 (i) SEQUENCE CHARACTERISTICS:

52

amino acids

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

10

15

20

25

30

35

53

		245		250		255
	Ala Val Arg Val Ala Ile Leu Val Thr Ser Ile Phe Phe Leu Cys Trp					
		260		265		270
5	Ser Pro Tyr His Ile Val Ile Phe Leu Asp Thr Leu Ala Arg Leu Lys					
		275		280		285
	Ala Val Asp Asn Thr Cys Lys Leu Asn Gly Ser Leu Pro Val Ala Ile					
		290		295		300
	Thr Met Cys Glu Phe Leu Gly Leu Ala His Cys Cys Leu Asn Pro Met					
		305		310		315
10	Leu Tyr Thr Phe Ala Gly Val Lys Phe Arg Ser Asp Leu Ser Arg Leu					
		325		330		335
	Leu Thr Lys Leu Gly Cys Thr Gly Pro Ala Ser Leu Cys Gln Leu Phe					
		340		345		350
15	Pro Ser Trp Arg Arg Ser Ser Leu Ser Glu Ser Glu Asn Ala Thr Ser					
		355		360		365
	Leu Thr Thr Phe					
		370				

(68) INFORMATION FOR SEQ ID NO:67:

- 20 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

CAAAGCTTGA AAGCTGCACG GTGCAGAGAC

30

(69) INFORMATION FOR SEQ ID NO:68:

- 30 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

35 GCGGATCCCC AGTCACACCC TGGCTGGGCC

30

(70) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1128 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

5

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

ATGGATGTGA CTTCCCAAGC CCGGGGCGTG GGCCTGGAGA TGTACCCAGG CACCGCGCAG 60
 CCTGCGGCCC CCAACACCAC CTCCCCGAG CTCAACCTGT CCCACCCGCT CCTGGGCACC 120
 10 GCCCTGGCCA ATGGGACAGG TGAGCTCTCG GAGCACCAGC AGTACGTGAT CGGCCTGTTC 180
 CTCTCGTGCC TCTACACCAT CTTCTCTTTC CCCATCGGCT TTGTGGGCAA CATCCTGATC 240
 CTGGTGGTGA ACATCAGCTT CCGCGAGAAG ATGACCATCC CCGACCTGTA CTTTCATCAAC 300
 CTGGCGGTGG CGGACCTCAT CCTGGTGGCC GACTCCCTCA TTGAGGTGTT CAACCTGCAC 360
 GAGCGGTACT ACGACATCGC CGTCCTGTGC ACCTTCATGT CGCTCTTCCT GCAGGTCAAC 420
 15 ATGTACAGCA GCGTCTTCTT CCTCACCTGG ATGAGCTTCG ACCGCTACAT CGCCCTGGCC 480
 AGGGCCATGC GCTGCAGCCT GTTCCGCACC AAGCACCACG CCCGGCTGAG CTGTGGCCTC 540
 ATCTGGATGG CATCCGTGTC AGCCACGCTG GTGCCCTTCA CCGCCGTGCA CCTGCAGCAC 600
 ACCGACGAGG CTGCTTCTG TTTCGCGGAT GTCCGGGAGG TGCAGTGGCT CGAGGTCACG 660
 CTGGGCTTCA TCGTGCCCTT CGCCATCATC GGCCTGTGCT ACTCCCTCAT TGTCCGGGTG 720
 20 CTGGTCAGGG CGCACC GGCA CCGTGGGCTG CGGCCCCGGC GGCAGAAGGC GCTCCGCATG 780
 ATCCTCGCGG TGGTGCTGGT CTTCTTCGTC TGCTGGCTGC CGGAGAACGT CTTTCATCAGC 840
 GTGCACCTCC TGCAGCGGAC GCAGCCTGGG GCCGCTCCCT GCAAGCAGTC TTTCCGCCAT 900
 GCCCACCCCC TCACGGGCCA CATTGTCAAC CTCACCGCCT TCTCCAACAG CTGCCTAAAC 960
 CCCCTCATCT ACAGCTTCTT CGGGGAGACC TTCAGGGACA AGCTGAGGCT GTACATTGAG 1020
 25 CAGAAAACAA ATTTGCCGGC CCTGAACCGC TTCTGTCACG CTGCCCTGAA GGCCGTCATT 1080
 CCAGACAGCA CCGAGCAGTC GGATGTGAGG TTCAGCAGTG CCGTGTAG 1128

(71) INFORMATION FOR SEQ ID NO:70:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 375 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

30

55

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

	Met	Asp	Val	Thr	Ser	Gln	Ala	Arg	Gly	Val	Gly	Leu	Glu	Met	Tyr	Pro	
	1				5					10					15		
5	Gly	Thr	Ala	Gln	Pro	Ala	Ala	Pro	Asn	Thr	Thr	Ser	Pro	Glu	Leu	Asn	
				20					25					30			
	Leu	Ser	His	Pro	Leu	Leu	Gly	Thr	Ala	Leu	Ala	Asn	Gly	Thr	Gly	Glu	
			35					40					45				
10	Leu	Ser	Glu	His	Gln	Gln	Tyr	Val	Ile	Gly	Leu	Phe	Leu	Ser	Cys	Leu	
		50					55					60					
	Tyr	Thr	Ile	Phe	Leu	Phe	Pro	Ile	Gly	Phe	Val	Gly	Asn	Ile	Leu	Ile	
	65					70					75					80	
	Leu	Val	Val	Asn	Ile	Ser	Phe	Arg	Glu	Lys	Met	Thr	Ile	Pro	Asp	Leu	
				85						90					95		
15	Tyr	Phe	Ile	Asn	Leu	Ala	Val	Ala	Asp	Leu	Ile	Leu	Val	Ala	Asp	Ser	
				100					105					110			
	Leu	Ile	Glu	Val	Phe	Asn	Leu	His	Glu	Arg	Tyr	Tyr	Asp	Ile	Ala	Val	
			115					120					125				
20	Leu	Cys	Thr	Phe	Met	Ser	Leu	Phe	Leu	Gln	Val	Asn	Met	Tyr	Ser	Ser	
		130					135						140				
	Val	Phe	Phe	Leu	Thr	Trp	Met	Ser	Phe	Asp	Arg	Tyr	Ile	Ala	Leu	Ala	
	145					150					155					160	
	Arg	Ala	Met	Arg	Cys	Ser	Leu	Phe	Arg	Thr	Lys	His	His	Ala	Arg	Leu	
				165						170					175		
25	Ser	Cys	Gly	Leu	Ile	Trp	Met	Ala	Ser	Val	Ser	Ala	Thr	Leu	Val	Pro	
				180				185						190			
	Phe	Thr	Ala	Val	His	Leu	Gln	His	Thr	Asp	Glu	Ala	Cys	Phe	Cys	Phe	
			195					200					205				
30	Ala	Asp	Val	Arg	Glu	Val	Gln	Trp	Leu	Glu	Val	Thr	Leu	Gly	Phe	Ile	
		210					215					220					
	Val	Pro	Phe	Ala	Ile	Ile	Gly	Leu	Cys	Tyr	Ser	Leu	Ile	Val	Arg	Val	
	225					230					235					240	
	Leu	Val	Arg	Ala	His	Arg	His	Arg	Gly	Leu	Arg	Pro	Arg	Arg	Gln	Lys	
				245						250					255		
35	Ala	Leu	Arg	Met	Ile	Leu	Ala	Val	Val	Leu	Val	Phe	Phe	Val	Cys	Trp	
				260					265					270			

Leu Pro Glu Asn Val Phe Ile Ser Val His Leu Leu Gln Arg Thr Gln
 275 280 285

Pro Gly Ala Ala Pro Cys Lys Gln Ser Phe Arg His Ala His Pro Leu
 290 295 300

5 Thr Gly His Ile Val Asn Leu Thr Ala Phe Ser Asn Ser Cys Leu Asn
 305 310 315 320

Pro Leu Ile Tyr Ser Phe Leu Gly Glu Thr Phe Arg Asp Lys Leu Arg
 325 330 335

10 Leu Tyr Ile Glu Gln Lys Thr Asn Leu Pro Ala Leu Asn Arg Phe Cys
 340 345 350

His Ala Ala Leu Lys Ala Val Ile Pro Asp Ser Thr Glu Gln Ser Asp
 355 360 365

Val Arg Phe Ser Ser Ala Val
 370 375

15 (72) INFORMATION FOR SEQ ID NO:71:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

20 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

ACAGAATTCC TGTGTGGTTT TACCGCCCAG

30

(73) INFORMATION FOR SEQ ID NO:72:

25 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

30 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

CTCGGATCCA GGCAGAAGAG TCGCCTATGG

30

(74) INFORMATION FOR SEQ ID NO:73:

35 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1137 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

```
ATGGACCTGG GGAAACCAAT GAAAAGCGTG CTGGTGGTGG CTCTCCTTGT CATTTTCCAG      60
5  GTATGCCTGT GTCAAGATGA GGTACGGAC GATTACATCG GAGACAACAC CACAGTGGAC      120
TACACTTTGT TCGAGTCTTT GTGCTCCAAG AAGGACGTGC GGAACTTTAA AGCCTGGTTC      180
CTCCCTATCA TGTACTCCAT CATTTGTTTC GTGGGCCTAC TGGGCAATGG GCTGGTCGTG      240
TTGACCTATA TCTATTTCAA GAGGCTCAAG ACCATGACCG ATACCTACCT GCTCAACCTG      300
GCGGTGGCAG ACATCCTCTT CCTCCTGACC CTTCCCTTCT GGGCCTACAG CGCGGCCAAG      360
10 TCCTGGGTCT TCGGTGTCCA CTTTGTGCAAG CTCATCTTTG CCATCTACAA GATGAGCTTC      420
TTCAGTGGCA TGCTCCTACT TCTTGTGATC AGCATTGACC GCTACGTGGC CATCGTCCAG      480
GCTGTCTCAG CTCACCGCCA CCGTGCCCGC GTCCTTCTCA TCAGCAAGCT GTCCTGTGTG      540
GGCATCTGGA TACTAGCCAC AGTGCTCTCC ATCCCAGAGC TCCTGTACAG TGACCTCCAG      600
AGGAGCAGCA GTGAGCAAGC GATGCGATGC TCTCTCATCA CAGAGCATGT GGAGGCCTTT      660
15 ATCACCATCC AGGTGGCCCA GATGGTGATC GGCTTTCTGG TCCCCCTGCT GGCCATGAGC      720
TTCTGTTACC TTGTCATCAT CCGCACCTG CTCCAGGCAC GCAACTTTGA GCGCAACAAG      780
GCCATCAAGG TGATCATCGC TGTGGTCGTG GTCTTCATAG TCTTCCAGCT GCCCTACAAT      840
GGGGTGGTCC TGGCCCAGAC GGTGGCCAAC TTCAACATCA CCAGTAGCAC CTGTGAGCTC      900
AGTAAGCAAC TCAACATCGC CTACGACGTC ACCTACAGCC TGGCCTGCGT CCGCTGCTGC      960
20 GTCAACCCTT TCTGTACGC CTTATCGGC GTCAAGTTCC GCAACGATCT CTTCAAGCTC      1020
TTCAAGGACC TGGGCTGCCT CAGCCAGGAG CAGCTCCGGC AGTGGTCTTC CTGTGCGCAC      1080
ATCCGGCGCT CCTCCATGAG TGTGGAGGCC GAGACCACCA CCACCTTCTC CCCATAG      1137
```

(75) INFORMATION FOR SEQ ID NO:74:

25 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 378 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

	Met	Asp	Leu	Gly	Lys	Pro	Met	Lys	Ser	Val	Leu	Val	Val	Ala	Leu	Leu	
	1				5					10					15		
	Val	Ile	Phe	Gln	Val	Cys	Leu	Cys	Cln	Asp	Glu	Val	Thr	Asp	Asp	Tyr	
				20					25					30			
5	Ile	Gly	Asp	Asn	Thr	Thr	Val	Asp	Tyr	Thr	Leu	Phe	Glu	Ser	Leu	Cys	
			35					40					45				
	Ser	Lys	Lys	Asp	Val	Arg	Asn	Phe	Lys	Ala	Trp	Phe	Leu	Pro	Ile	Met	
		50					55					60					
10	Tyr	Ser	Ile	Ile	Cys	Phe	Val	Gly	Leu	Leu	Gly	Asn	Gly	Leu	Val	Val	
	65					70					75					80	
	Leu	Thr	Tyr	Ile	Tyr	Phe	Lys	Arg	Leu	Lys	Thr	Met	Thr	Asp	Thr	Tyr	
				85						90					95		
	Leu	Leu	Asn	Leu	Ala	Val	Ala	Asp	Ile	Leu	Phe	Leu	Leu	Thr	Leu	Pro	
				100					105						110		
15	Phe	Trp	Ala	Tyr	Ser	Ala	Ala	Lys	Ser	Trp	Val	Phe	Gly	Val	His	Phe	
			115					120					125				
	Cys	Lys	Leu	Ile	Phe	Ala	Ile	Tyr	Lys	Met	Ser	Phe	Phe	Ser	Gly	Met	
		130					135					140					
20	Leu	Leu	Leu	Leu	Cys	Ile	Ser	Ile	Asp	Arg	Tyr	Val	Ala	Ile	Val	Gln	
	145					150					155				160		
	Ala	Val	Ser	Ala	His	Arg	His	Arg	Ala	Arg	Val	Leu	Leu	Ile	Ser	Lys	
				165					170						175		
	Leu	Ser	Cys	Val	Gly	Ile	Trp	Ile	Leu	Ala	Thr	Val	Leu	Ser	Ile	Pro	
				180				185					190				
25	Glu	Leu	Leu	Tyr	Ser	Asp	Leu	Gln	Arg	Ser	Ser	Ser	Glu	Gln	Ala	Met	
		195					200						205				
	Arg	Cys	Ser	Leu	Ile	Thr	Glu	His	Val	Glu	Ala	Phe	Ile	Thr	Ile	Gln	
		210				215					220						
30	Val	Ala	Gln	Met	Val	Ile	Gly	Phe	Leu	Val	Pro	Leu	Leu	Ala	Met	Ser	
	225					230					235				240		
	Phe	Cys	Tyr	Leu	Val	Ile	Ile	Arg	Thr	Leu	Leu	Gln	Ala	Arg	Asn	Phe	
				245						250				255			
	Glu	Arg	Asn	Lys	Ala	Ile	Lys	Val	Ile	Ile	Ala	Val	Val	Val	Val	Phe	
				260				265					270				
35	Ile	Val	Phe	Gln	Leu	Pro	Tyr	Asn	Gly	Val	Val	Leu	Ala	Gln	Thr	Val	
		275					280						285				

59

Ala Asn Phe Asn Ile Thr Ser Ser Thr Cys Glu Leu Ser Lys Gln Leu
 290 295 300

Asn Ile Ala Tyr Asp Val Thr Tyr Ser Leu Ala Cys Val Arg Cys Cys
 305 310 315 320

5 Val Asn Pro Phe Leu Tyr Ala Phe Ile Gly Val Lys Phe Arg Asn Asp
 325 330 335

Leu Phe Lys Leu Phe Lys Asp Leu Gly Cys Leu Ser Gln Glu Gln Leu
 340 345 350

10 Arg Gln Trp Ser Ser Cys Arg His Ile Arg Arg Ser Ser Met Ser Val
 355 360 365

Glu Ala Glu Thr Thr Thr Thr Phe Ser Pro
 370 375

(76) INFORMATION FOR SEQ ID NO:75:

15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 32 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

CTGGAATTCA CCTGGACCAC CACCAATGGA TA

32

(77) INFORMATION FOR SEQ ID NO:76:

25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

30 CTCGGATCCT GCAAAGTTTG TCATACAGTT

30

(78) INFORMATION FOR SEQ ID NO:77:

35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1085 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

ATGGATATAC AAATGGCAAA CAATTTTACT CCGCCCTCTG CAACTCCTCA GGGAAATGAC 60
 TGTGACCTCT ATGCACATCA CAGCACGGCC AGGATAGTAA TGCCTCTGCA TTACAGCCTC 120
 GTCTTCATCA TTGGGCTCGT GGGAAACTTA CTAGCCTTGG TCGTCATTGT TCAAAACAGG 180
 5 AAAAAAATCA ACTCTACCAC CCTCTATTCA ACAAATTGG TGATTCTGA TATACTTTTT 240
 ACCACGGCTT TGCCTACACG AATAGCCTAC TATGCAATGG GCTTTGACTG GAGAATCGGA 300
 GATGCCTTGT GTAGGATAAC TGCGCTAGTG TTTTACATCA ACACATATGC AGGTGTGAAC 360
 TTTATGACCT GCCTGAGTAT TGACCGCTTC ATTGCTGTGG TGCACCCTCT ACGCTACAAC 420
 AAGATAAAAA GGATTGAACA TGCAAAGGC GTGTGCATAT TTGTCTGGAT TCTAGTATTT 480
 10 GCTCAGACAC TCCCACTCCT CATCAACCCT ATGTCAAAGC AGGAGGCTGA AAGGATTACA 540
 TGCATGGAGT ATCCAAACTT TGAAGAACT AAATCTCTTC CCTGGATTCT GCTTGGGGCA 600
 TGTTTCATAG GATATGTACT TCCACTTATA ATCATTCTCA TCTGCTATTC TCAGATCTGC 660
 TGCAAACCTCT TCAGAACTGC CAAACAAAAC CCACTCACTG AGAAATCTGG TGTAACAAA 720
 AAGGCTCTCA ACACAATTAT TCTTATTATT GTTGTGTTTG TTCTCTGTTT CACACCTTAC 780
 15 CATGTTGCAA TTATTCAACA TATGATTAAG AAGCTTCGTT TCTCTAATTT CCTGGAATGT 840
 AGCCAAAGAC ATTCGTTCCA GATTTCTCTG CACTTTACAG TATGCCTGAT GAACTTCAAT 900
 TGCTGCATGG ACCCTTTTAT CTACTTCTTT GCATGTAAAG GGTATAAGAG AAAGGTTATG 960
 AGGATGCTGA AACGGCAAGT CAGTGTATCG ATTTCTAGTG CTGTGAAGTC AGCCCCTGAA 1020
 GAAAATTCAC GTGAAATGAC AGAAACGCAG ATGATGATAC ATTCCAAGTC TTCAAATGGA 1080
 20 AAGTGA 1086

(79) INFORMATION FOR SEQ ID NO:78:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 361 amino acids
 (B) TYPE: amino acid
 25 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

30 Met Asp Ile Gln Met Ala Asn Asn Phe Thr Pro Pro Ser Ala Thr Pro
 1 5 10 15

61

	Gln	Gly	Asn	Asp	Cys	Asp	Leu	Tyr	Ala	His	His	Ser	Thr	Ala	Arg	Ile	
				20					25					30			
	Val	Met	Pro	Leu	His	Tyr	Ser	Leu	Val	Phe	Ile	Ile	Gly	Leu	Val	Gly	
			35					40					45				
5	Asn	Leu	Leu	Ala	Leu	Val	Val	Ile	Val	Gln	Asn	Arg	Lys	Lys	Ile	Asn	
		50					55					60					
	Ser	Thr	Thr	Leu	Tyr	Ser	Thr	Asn	Leu	Val	Ile	Ser	Asp	Ile	Leu	Phe	
	65					70					75					80	
10	Thr	Thr	Ala	Leu	Pro	Thr	Arg	Ile	Ala	Tyr	Tyr	Ala	Met	Gly	Phe	Asp	
					85					90					95		
	Trp	Arg	Ile	Gly	Asp	Ala	Leu	Cys	Arg	Ile	Thr	Ala	Leu	Val	Phe	Tyr	
				100					105					110			
	Ile	Asn	Thr	Tyr	Ala	Gly	Val	Asn	Phe	Met	Thr	Cys	Leu	Ser	Ile	Asp	
			115					120					125				
15	Arg	Phe	Ile	Ala	Val	Val	His	Pro	Leu	Arg	Tyr	Asn	Lys	Ile	Lys	Arg	
		130					135					140					
	Ile	Glu	His	Ala	Lys	Gly	Val	Cys	Ile	Phe	Val	Trp	Ile	Leu	Val	Phe	
	145					150				155						160	
20	Ala	Gln	Thr	Leu	Pro	Leu	Leu	Ile	Asn	Pro	Met	Ser	Lys	Gln	Glu	Ala	
					165				170						175		
	Glu	Arg	Ile	Thr	Cys	Met	Glu	Tyr	Pro	Asn	Phe	Glu	Glu	Thr	Lys	Ser	
				180					185					190			
	Leu	Pro	Trp	Ile	Leu	Leu	Gly	Ala	Cys	Phe	Ile	Gly	Tyr	Val	Leu	Pro	
			195					200					205				
25	Leu	Ile	Ile	Ile	Leu	Ile	Cys	Tyr	Ser	Gln	Ile	Cys	Cys	Lys	Leu	Phe	
		210					215					220					
	Arg	Thr	Ala	Lys	Gln	Asn	Pro	Leu	Thr	Glu	Lys	Ser	Gly	Val	Asn	Lys	
	225					230					235					240	
30	Lys	Ala	Leu	Asn	Thr	Ile	Ile	Leu	Ile	Ile	Val	Val	Phe	Val	Leu	Cys	
					245					250					255		
	Phe	Thr	Pro	Tyr	His	Val	Ala	Ile	Ile	Gln	His	Met	Ile	Lys	Lys	Leu	
				260					265					270			
	Arg	Phe	Ser	Asn	Phe	Leu	Glu	Cys	Ser	Gln	Arg	His	Ser	Phe	Gln	Ile	
			275					280					285				
35	Ser	Leu	His	Phe	Thr	Val	Cys	Leu	Met	Asn	Phe	Asn	Cys	Cys	Met	Asp	
		290					295					300					
	Pro	Phe	Ile	Tyr	Phe	Phe	Ala	Cys	Lys	Gly	Tyr	Lys	Arg	Lys	Val	Met	

```

305           310           315           320
Arg Met Leu Lys Arg Gln Val Ser Val Ser Ile Ser Ser Ala Val Lys
           325           330           335
Ser Ala Pro Glu Glu Asn Ser Arg Glu Met Thr Glu Thr Gln Met Met
           340           345           350
Ile His Ser Lys Ser Ser Asn Gly Lys
           355           360

```

(80) INFORMATION FOR SEQ ID NO:79:

- 10 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- 15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

CTGGAATTCT CCTGCTCATC CAGCCATGCG G 31

(81) INFORMATION FOR SEQ ID NO:80:

- 20
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

25 CCTGGATCCC CACCCCTACT GGGGCCTCAG 30

(82) INFORMATION FOR SEQ ID NO:81:

- 30
- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1446 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

ATGCGGTGGC TGTGGCCCCT GGCTGTCTCT CTTGCTGTGA TTTTGGCTGT GGGGCTAAGC 60

35 AGGGTCTCTG GGGGTGCCCC CCTGCACCTG GGCAGGCACA GAGCCGAGAC CCAGGAGCAG 120

CAGAGCCGAT CCAAGAGGGG CACCGAGGAT GAGGAGGCCA AGGGCGTGCA GCAGTATGTG 180
CCTGAGGAGT GGGCGGAGTA CCCCCGCCCC ATTCACCCTG CTGGCCTGCA GCCAACCAAG 240
CCCTTGGTGG CCACCAGCCC TAACCCCGAC AAGGATGGGG GCACCCCA GAAGTGGGCAG 300
GAACTGAGGG GCAATCTGAC AGGGGCACCA GGGCAGAGGC TACAGATCCA GAACCCCTG 360
5 TATCCGGTGA CCGAGAGCTC CTACAGTGCC TATGCCATCA TGCTTCTGGC GCTGGTGGTG 420
TTTGC GGTTGG GCATTGTGGG CAACCTGTCTG GTCATGTGCA TCGTGTGGCA CAGCTACTAC 480
CTGAAGAGCG CCTGGAAGTC CATCCTTGCC AGCCTGGCCC TCTGGGATTT TCTGGTCCTC 540
TTTTTCTGCC TCCCTATTGT CATCTTCAAC GAGATCACCA AGCAGAGGCT ACTGGGTGAC 600
GTTTCTTGTC GTGCCGTGCC CTTCATGGAG GTCTCCTCTC TGGGAGTCAC GACTTTCAGC 660
10 CTCTGTGCCC TGGGCATTGA CCGCTTCCAC GTGGCCACCA GCACCCTGCC CAAGGTGAGG 720
CCCATCGAGC GGTGCCAATC CATCCTGGCC AAGTTGGCTG TCATCTGGGT GGGCTCCATG 780
ACGCTGGCTG TGCCTGAGCT CCTGCTGTGG CAGCTGGCAC AGGAGCCTGC CCCCACCATG 840
GGCACCTGG ACTCATGCAT CATGAAACCC TCAGCCAGCC TGCCCGAGTC CCTGTATTCA 900
CTGGTGATGA CCTACCAGAA CGCCCGCATG TGGTGGTACT TTGGCTGCTA CTTCTGCCTG 960
15 CCCATCCTCT TCACAGTCAC CTGCCAGCTG GTGACATGGC GGGTGCAGAG CCCTCCAGGG 1020
AGGAAGTCAG AGTGCAGGGC CAGCAAGCAC GAGCAGTGTG AGAGCCAGCT CAACAGCACC 1080
GTGGTGGGCC TGACCGTGGT CTACGCCTTC TGCACCCTCC CAGAGAACGT CTGCAACATC 1140
GTGGTGGCCT ACCTCTCCAC CGAGCTGACC CGCCAGACCC TGGACCTCCT GGGCCTCATC 1200
AACCAGTTCT CCACCTTCTT CAAGGGCGCC ATCACCCAG TGCTGCTCCT TTGCATCTGC 1260
20 AGGCCGCTGG GCCAGGCCTT CCTGGACTGC TGCTGCTGCT GCTGCTGTGA GGAGTGCGGC 1320
GGGGCTTCGG AGGCCTCTGC TGCCAATGGG TCGGACAACA AGCTCAAGAC CGAGGTGTCC 1380
TCTTCCATCT ACTTCCACAA GCCCAGGGAG TCACCCCCAC TCCTGCCCCT GGGCACACCT 1440
TGCTGA 1446

(83) INFORMATION FOR SEQ ID NO:82:

- 25 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 481 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: not relevant
- 30 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

	Met	Arg	Trp	Leu	Trp	Pro	Leu	Ala	Val	Ser	Leu	Ala	Val	Ile	Leu	Ala
	1				5						10				15	
5	Val	Gly	Leu	Ser	Arg	Val	Ser	Gly	Gly	Ala	Pro	Leu	His	Leu	Gly	Arg
			20						25					30		
	His	Arg	Ala	Glu	Thr	Gln	Glu	Gln	Gln	Ser	Arg	Ser	Lys	Arg	Gly	Thr
			35					40					45			
	Glu	Asp	Glu	Glu	Ala	Lys	Gly	Val	Gln	Gln	Tyr	Val	Pro	Glu	Glu	Trp
		50					55					60				
10	Ala	Glu	Tyr	Pro	Arg	Pro	Ile	His	Pro	Ala	Gly	Leu	Gln	Pro	Thr	Lys
	65					70					75					80
	Pro	Leu	Val	Ala	Thr	Ser	Pro	Asn	Pro	Asp	Lys	Asp	Gly	Gly	Thr	Pro
					85					90					95	
	Asp	Ser	Gly	Gln	Glu	Leu	Arg	Gly	Asn	Leu	Thr	Gly	Ala	Pro	Gly	Gln
15				100					105					110		
	Arg	Leu	Gln	Ile	Gln	Asn	Pro	Leu	Tyr	Pro	Val	Thr	Glu	Ser	Ser	Tyr
			115					120					125			
	Ser	Ala	Tyr	Ala	Ile	Met	Leu	Leu	Ala	Leu	Val	Val	Phe	Ala	Val	Gly
		130					135					140				
20	Ile	Val	Gly	Asn	Leu	Ser	Val	Met	Cys	Ile	Val	Trp	His	Ser	Tyr	Tyr
	145				150						155					160
	Leu	Lys	Ser	Ala	Trp	Asn	Ser	Ile	Leu	Ala	Ser	Leu	Ala	Leu	Trp	Asp
					165					170					175	
	Phe	Leu	Val	Leu	Phe	Phe	Cys	Leu	Pro	Ile	Val	Ile	Phe	Asn	Glu	Ile
25				180					185					190		
	Thr	Lys	Gln	Arg	Leu	Leu	Gly	Asp	Val	Ser	Cys	Arg	Ala	Val	Pro	Phe
			195					200					205			
	Met	Glu	Val	Ser	Ser	Leu	Gly	Val	Thr	Thr	Phe	Ser	Leu	Cys	Ala	Leu
		210					215					220				
30	Gly	Ile	Asp	Arg	Phe	His	Val	Ala	Thr	Ser	Thr	Leu	Pro	Lys	Val	Arg
	225					230					235					240
	Pro	Ile	Glu	Arg	Cys	Gln	Ser	Ile	Leu	Ala	Lys	Leu	Ala	Val	Ile	Trp
					245					250					255	
	Val	Gly	Ser	Met	Thr	Leu	Ala	Val	Pro	Glu	Leu	Leu	Leu	Trp	Gln	Leu
35				260					265							
	Ala	Gln	Glu	Pro	Ala	Pro	Thr	Met	Gly	Thr	Leu	Asp	Ser	Cys	Ile	Met

65

	275	280	285
	Lys Pro Ser Ala Ser Leu	Pro Glu Ser Leu Tyr	Ser Leu Val Met Thr
	290	295	300
5	Tyr Gln Asn Ala Arg Met Trp Trp Tyr Phe Gly Cys Tyr Phe Cys Leu		
	305	310	315 320
	Pro Ile Leu Phe Thr Val Thr Cys Gln Leu Val Thr Trp Arg Val Arg		
		325	330 335
	Gly Pro Pro Gly Arg Lys Ser Glu Cys Arg Ala Ser Lys His Glu Gln		
		340	345 350
10	Cys Glu Ser Gln Leu Asn Ser Thr Val Val Gly Leu Thr Val Val Tyr		
		355	360 365
	Ala Phe Cys Thr Leu Pro Glu Asn Val Cys Asn Ile Val Val Ala Tyr		
		370	375 380
15	Leu Ser Thr Glu Leu Thr Arg Gln Thr Leu Asp Leu Leu Gly Leu Ile		
		385	390 395 400
	Asn Gln Phe Ser Thr Phe Phe Lys Gly Ala Ile Thr Pro Val Leu Leu		
		405	410 415
	Leu Cys Ile Cys Arg Pro Leu Gly Gln Ala Phe Leu Asp Cys Cys Cys		
		420	425 430
20	Cys Cys Cys Cys Glu Glu Cys Gly Gly Ala Ser Glu Ala Ser Ala Ala		
		435	440 445
	Asn Gly Ser Asp Asn Lys Leu Lys Thr Glu Val Ser Ser Ser Ile Tyr		
		450	455 460
25	Phe His Lys Pro Arg Glu Ser Pro Pro Leu Leu Pro Leu Gly Thr Pro		
		465	470 475 480
	Cys		

(84) INFORMATION FOR SEQ ID NO:83:

30 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

ATGTGGAACG CGACGCCAG CG

(85) INFORMATION FOR SEQ ID NO:84:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

TCATGTATTA ATACTAGATT CT

22

10 (86) INFORMATION FOR SEQ ID NO:85:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 38 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

TACCATGTGG AACGCGACGC CCAGCGAAGA GCCGGGGT

38

(87) INFORMATION FOR SEQ ID NO:86:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

CGGAATTCAT GTATTAATAC TAGATTCTGT CCAGGCCCG

39

(88) INFORMATION FOR SEQ ID NO:87:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1101 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

35

ATGTGGAACG CGACGCCCAG CGAAGAGCCG GGGTTCAACC TCACACTGGC CGACCTGGAC 60
 TGGGATGCTT CCCCCGGCAA CGACTCGCTG GGCGACGAGC TGCTGCAGCT CTTCCCCGCG 120
 CCGCTGCTGG CGGGCGTCAC AGCCACCTGC GTGGCACTCT TCGTGGTGGG TATCGCTGGC 180
 AACCTGCTCA CCATGCTGGT GGTGTCGCGC TTCCGCGAGC TGCGCACCAC CACCAACCTC 240
 5 TACCTGTCCA GCATGGCCTT CTCCGATCTG CTCATCTTCC TCTGCATGCC CCTGGACCTC 300
 GTTCGCCTCT GGCAGTACCG GCCCTGGAAC TTCGGCGACC TCCTCTGCAA ACTCTTCCAA 360
 TTCGTCAGTG AGAGCTGCAC CTACGCCACG GTGCTCACCA TCACAGCGCT GAGCGTCGAG 420
 CGCTACTTCG CCATCTGCTT CCCACTCCGG GCCAAGGTGG TGGTCACCAA GGGGCGGGTG 480
 AAGCTGGTCA TCTTCGTCAT CTGGGCCGTG GCCTTCTGCA GCGCCGGGCC CATCTTCGTG 540
 10 CTAGTCGGGG TGGAGCACGA GAACGGCACC GACCCTTGGG ACACCAACGA GTGCCGCCCC 600
 ACCGAGTTTG CGGTGCGCTC TGGACTGCTC ACGGTCATGG TGTGGGTGTC CAGCATCTTC 660
 TTCTTCCTTC CTGTCTTCTG TCTCACGGTC CTCTACAGTC TCATCGGCAG GAAGCTGTGG 720
 CGGAGGAGGC GCGGCGATGC TGTGCTGGGT GCCTCGCTCA GGGACCAGAA CCACAAGCAA 780
 ACCGTGAAAA TGCTGGCTGT AGTGGTGTTT GCCTTCATCC TCTGCTGGCT CCCCTTCCAC 840
 15 GTAGGGCGAT ATTTATTTTC CAAATCCTTT GAGCCTGGCT CTTGGAGAT TGCTCAGATC 900
 AGCCAGTACT GCAACCTCGT GTCCTTTGTC CTCTTCTACC TCAGTGCTGC CATCAACCCC 960
 ATTCTGTACA ACATCATGTC CAAGAAGTAC CGGGTGGCAG TGTTCAGACT TCTGGGATTC 1020
 GAACCCTTCT CCCAGAGAAA GCTCTCCACT CTGAAAGATG AAAGTTCTCG GGCCTGGACA 1080
 GAATCTAGTA TTAATACATG A 1101

20 (89) INFORMATION FOR SEQ ID NO:88:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 366 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

25 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

Met Trp Asn Ala Thr Pro Ser Glu Glu Pro Gly Phe Asn Leu Thr Leu
 1 5 10 15

30 Ala Asp Leu Asp Trp Asp Ala Ser Pro Gly Asn Asp Ser Leu Gly Asp
 20 25 30

Glu Leu Leu Gln Leu Phe Pro Ala Pro Leu Leu Ala Gly Val Thr Ala
 35 40 45
 Thr Cys Val Ala Leu Phe Val Val Gly Ile Ala Gly Asn Leu Leu Thr
 50 55 60
 5 Met Leu Val Val Ser Arg Phe Arg Glu Leu Arg Thr Thr Thr Asn Leu
 65 70 75 80
 Tyr Leu Ser Ser Met Ala Phe Ser Asp Leu Leu Ile Phe Leu Cys Met
 85 90 95
 10 Pro Leu Asp Leu Val Arg Leu Trp Gln Tyr Arg Pro Trp Asn Phe Gly
 100 105 110
 Asp Leu Leu Cys Lys Leu Phe Gln Phe Val Ser Glu Ser Cys Thr Tyr
 115 120 125
 Ala Thr Val Leu Thr Ile Thr Ala Leu Ser Val Glu Arg Tyr Phe Ala
 130 135 140
 15 Ile Cys Phe Pro Leu Arg Ala Lys Val Val Val Thr Lys Gly Arg Val
 145 150 155 160
 Lys Leu Val Ile Phe Val Ile Trp Ala Val Ala Phe Cys Ser Ala Gly
 165 170 175
 20 Pro Ile Phe Val Leu Val Gly Val Glu His Glu Asn Gly Thr Asp Pro
 180 185 190
 Trp Asp Thr Asn Glu Cys Arg Pro Thr Glu Phe Ala Val Arg Ser Gly
 195 200 205
 Leu Leu Thr Val Met Val Trp Val Ser Ser Ile Phe Phe Phe Leu Pro
 210 215 220
 25 Val Phe Cys Leu Thr Val Leu Tyr Ser Leu Ile Gly Arg Lys Leu Trp
 225 230 235 240
 Arg Arg Arg Arg Gly Asp Ala Val Val Gly Ala Ser Leu Arg Asp Gln
 245 250 255
 30 Asn His Lys Gln Thr Val Lys Met Leu Ala Val Val Val Phe Ala Phe
 260 265 270
 Ile Leu Cys Trp Leu Pro Phe His Val Gly Arg Tyr Leu Phe Ser Lys
 275 280 285
 Ser Phe Glu Pro Gly Ser Leu Glu Ile Ala Gln Ile Ser Gln Tyr Cys
 290 295 300
 35 Asn Leu Val Ser Phe Val Leu Phe Tyr Leu Ser Ala Ala Ile Asn Pro
 305 310 315 320
 Ile Leu Tyr Asn Ile Met Ser Lys Lys Tyr Arg Val Ala Val Phe Arg

5

10

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

33

20

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

30

25

- 30

- (xi) SEQUENCE DESCRIPTION: SEO ID NO:91:

60

120

180

240

CTTGCGGGAC CCTCCTGGGA CTGCGCGGCG GCGCCGGGCC GTGACCCGGC TGCAGGCAGA 300
 GGGGCGGAGG CGTCGGCAGC CGGACCCCCG GGACCTCCAA CCAGGCCACC TGGCCCCTGG 360
 AGGTGGAAG GTGCTCGGGG TCAGGAGCCT TCTGAACTT TGGGGAGAGG GAACCCACG 420
 GCCCTCCAGC TCTTCCTTCA GATCTCAGAG GAGGAAGAGA AGGGTCCCAG AGGCGCTGGC 480
 5 ATTTCCGGGC GTAGCCAGGA GCAGAGTGTG AAGACAGTCC CCGGAGCCAG CGATCTTTTT 540
 TACTGGCCAA GGAGAGCCGG GAAACTCCAG GGTTCACACC ACAAGCCCCT GTCCAAGACG 600
 GCCAATGGAC TGGCGGGGCA CGAAGGGTGG ACAATTGCAC TCCCGGGCCG GGCCTGGCC 660
 CAGAATGGAT CCTTGGGTGA AGGAATCCAT GAGCCTGGGG GTCCCCGCCG GGGAAACAGC 720
 ACGAACCGGC GTGTGAGACT GAAGAACCCC TTCTACCCGC TGACCCAGGA GTCCTATGGA 780
 10 GCCTACGCGG TCATGTGTCT GTCCGTGGTG ATCTTCGGGA CCGGCATCAT TGGCAACCTG 840
 GCGGTGATGA GCATCGTGTG CCACAACCTAC TACATGCGGA GCATCTCCAA CTCCCTCTTG 900
 GCCAACCTGG CCTTCTGGGA CTTTCTCATC ATCTTCTTCT GCCTTCCGCT GGTCTCTTC 960
 CACGAGCTGA CCAAGAAGTG GCTGCTGGAG GACTTCTCCT GCAAGATCGT GCCCTATATA 1020
 GAGGTGCTT CTCTGGGAGT CACCACTTTC ACCTTATGTG CTCTGTGCAT AGACCGCTTC 1080
 15 CGTGCTGCCA CCAACGTACA GATGTACTAC GAAATGATCG AAACTGTTC CTCAACAACT 1140
 GCCAACTTG CTGTTATATG GGTGGGAGCT CTATTGTTAG CACTTCCAGA AGTTGTTCTC 1200
 CGCCAGCTGA GCAAGGAGGA TTTGGGGTTT AGTGGCCGAG CTCCGGCAGA AAGGTGCATT 1260
 ATTAAGATCT CTCCTGATTT ACCAGACACC ATCTATGTTT TAGCCCTCAC CTACGACAGT 1320
 GCGAGACTGT GGTGGTATTT TGGCTGTTAC TTTGTTTGC CCACGCTTTT CACCATCACC 1380
 20 TGCTCTCTAG TGA CTGCGAG GAAATCCGC AAAGCAGAGA AAGCCTGTAC CCGAGGGAAT 1440
 AAACGGCAGA TTCAACTAGA GAGTCAGATG AACTGTACAG TAGTGGCACT GACCATTTTA 1500
 TATGGATTTT GCATTATTCC TGAAAATATC TGCAACATTG TTA CTGCCTA CATGGCTACA 1560
 GGGGTTTCAC AGCAGACAAT GGACCTCCTT AATATCATCA GCCAGTTCCT TTTGTTCTTT 1620
 AAGTCCTGTG TCACCCAGT CCTCCTTTTC TGTCTCTGCA AACCTTCAG TCGGGCCTTC 1680
 25 ATGGAGTGCT GCTGCTGTTG CTGTGAGGAA TGCATTGAGA AGTCTTCAAC GGTGACCACT 1740
 GATGACAATG ACAACGAGTA CACCACGGAA CTCGAACTCT CGCCTTTCAG TACCATACGC 1800
 CGTGAAATGT CCACTTTTGC TTCTGTGCGA ACTCATTGCT GA 1842

(93) INFORMATION FOR SEQ ID NO:92:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 613 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

5

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:92:

	Met	Arg	Ala	Pro	Gly	Ala	Leu	Leu	Ala	Arg	Met	Ser	Arg	Leu	Leu	Leu	
	1				5					10					15		
10	Leu	Leu	Leu	Leu	Lys	Val	Ser	Ala	Ser	Ser	Ala	Leu	Gly	Val	Ala	Pro	
				20					25					30			
	Ala	Ser	Arg	Asn	Glu	Thr	Cys	Leu	Gly	Glu	Ser	Cys	Ala	Pro	Thr	Val	
			35					40					45				
15	Ile	Gln	Arg	Arg	Gly	Arg	Asp	Ala	Trp	Gly	Pro	Gly	Asn	Ser	Ala	Arg	
	50						55					60					
	Asp	Val	Leu	Arg	Ala	Arg	Ala	Pro	Arg	Glu	Glu	Gln	Gly	Ala	Ala	Phe	
	65					70				75						80	
	Leu	Ala	Gly	Pro	Ser	Trp	Asp	Leu	Pro	Ala	Ala	Pro	Gly	Arg	Asp	Pro	
					85				90						95		
20	Ala	Ala	Gly	Arg	Gly	Ala	Glu	Ala	Ser	Ala	Ala	Gly	Pro	Pro	Gly	Pro	
				100					105					110			
	Pro	Thr	Arg	Pro	Pro	Gly	Pro	Trp	Arg	Trp	Lys	Gly	Ala	Arg	Gly	Gln	
			115					120					125				
25	Glu	Pro	Ser	Glu	Thr	Leu	Gly	Arg	Gly	Asn	Pro	Thr	Ala	Leu	Gln	Leu	
	130						135					140					
	Phe	Leu	Gln	Ile	Ser	Glu	Glu	Glu	Glu	Lys	Gly	Pro	Arg	Gly	Ala	Gly	
	145					150					155					160	
	Ile	Ser	Gly	Arg	Ser	Gln	Glu	Gln	Ser	Val	Lys	Thr	Val	Pro	Gly	Ala	
				165						170					175		
30	Ser	Asp	Leu	Phe	Tyr	Trp	Pro	Arg	Arg	Ala	Gly	Lys	Leu	Gln	Gly	Ser	
			180						185					190			
	His	His	Lys	Pro	Leu	Ser	Lys	Thr	Ala	Asn	Gly	Leu	Ala	Gly	His	Glu	
			195					200					205				
35	Gly	Trp	Thr	Ile	Ala	Leu	Pro	Gly	Arg	Ala	Leu	Ala	Gln	Asn	Gly	Ser	
	210						215					220					
	Leu	Gly	Glu	Gly	Ile	His	Glu	Pro	Gly	Gly	Pro	Arg	Arg	Gly	Asn	Ser	
	225					230					235					240	

Thr Asn Arg Arg Val Arg Leu Lys Asn Pro Phe Tyr Pro Leu Thr Gln
 245 250 255
 Glu Ser Tyr Gly Ala Tyr Ala Val Met Cys Leu Ser Val Val Ile Phe
 260 265 270
 5 Gly Thr Gly Ile Ile Gly Asn Leu Ala Val Met Ser Ile Val Cys His
 275 280 285
 Asn Tyr Tyr Met Arg Ser Ile Ser Asn Ser Leu Leu Ala Asn Leu Ala
 290 295 300
 10 Phe Trp Asp Phe Leu Ile Ile Phe Phe Cys Leu Pro Leu Val Ile Phe
 305 310 315 320
 His Glu Leu Thr Lys Lys Trp Leu Leu Glu Asp Phe Ser Cys Lys Ile
 325 330 335
 Val Pro Tyr Ile Glu Val Ala Ser Leu Gly Val Thr Thr Phe Thr Leu
 340 345 350
 15 Cys Ala Leu Cys Ile Asp Arg Phe Arg Ala Ala Thr Asn Val Gln Met
 355 360 365
 Tyr Tyr Glu Met Ile Glu Asn Cys Ser Ser Thr Thr Ala Lys Leu Ala
 370 375 380
 20 Val Ile Trp Val Gly Ala Leu Leu Leu Ala Leu Pro Glu Val Val Leu
 385 390 395 400
 Arg Gln Leu Ser Lys Glu Asp Leu Gly Phe Ser Gly Arg Ala Pro Ala
 405 410 415
 Glu Arg Cys Ile Ile Lys Ile Ser Pro Asp Leu Pro Asp Thr Ile Tyr
 420 425 430
 25 Val Leu Ala Leu Thr Tyr Asp Ser Ala Arg Leu Trp Trp Tyr Phe Gly
 435 440 445
 Cys Tyr Phe Cys Leu Pro Thr Leu Phe Thr Ile Thr Cys Ser Leu Val
 450 455 460
 30 Thr Ala Arg Lys Ile Arg Lys Ala Glu Lys Ala Cys Thr Arg Gly Asn
 465 470 475 480
 Lys Arg Gln Ile Gln Leu Glu Ser Gln Met Asn Cys Thr Val Val Ala
 485 490 495
 Leu Thr Ile Leu Tyr Gly Phe Cys Ile Ile Pro Glu Asn Ile Cys Asn
 500 505 510
 35 Ile Val Thr Ala Tyr Met Ala Thr Gly Val Ser Gln Gln Thr Met Asp
 515 520 525
 Leu Leu Asn Ile Ile Ser Gln Phe Leu Leu Phe Phe Lys Ser Cys Val

73

530 535 540

Thr Pro Val Leu Leu Phe Cys Leu Cys Lys Pro Phe Ser Arg Ala Phe
 545 550 555 560

5 Met Glu Cys Cys Cys Cys Cys Cys Glu Glu Cys Ile Gln Lys Ser Ser
 565 570 575

Thr Val Thr Ser Asp Asp Asn Asp Asn Glu Tyr Thr Thr Glu Leu Glu
 580 585 590

Leu Ser Pro Phe Ser Thr Ile Arg Arg Glu Met Ser Thr Phe Ala Ser
 595 600 605

10 Val Gly Thr His Cys
 610

(94) INFORMATION FOR SEQ ID NO:93:

15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 34 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:93:

20 CAGAATTCAG AGAAAAAAG TGAATATGGT TTTT 34

(95) INFORMATION FOR SEQ ID NO:94:

25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 32 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:

TTGGATCCCT GGTGCATAAC AATTGAAAGA AT 32

30 (96) INFORMATION FOR SEQ ID NO:95:

35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1248 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:

ATGGTTTTGTG CTCACAGAAT GGATAACAGC AAGCCACATT TGATTATTCC TACACTTCTG 60
 GTGCCCTCC AAAACCGCAG CTGCACTGAA ACAGCCACAC CTCTGCCAAG CCAATACCTG 120
 ATGGAATTAA GTGAGGAGCA CAGTTGGATG AGCAACCAA CAGACCTTCA CTATGTGCTG 180
 5 AAACCCGGGG AAGTGGCCAC AGCCAGCATC TTCTTTGGGA TTCTGTGGTT GTTTTCTATC 240
 TTCGGCAATT CCCTGGTTTG TTTGGTCATC CATAGGAGTA GGAGGACTCA GTCTACCACC 300
 AACTACTTTG TGGTCTCCAT GGCATGTGCT GACCTTCTCA TCAGCGTTGC CAGCAGCCT 360
 TTCGTCCTGC TCCAGTTCAC CACTGGAAGG TGGACGCTGG GTAGTGCAAC GTGCAAGGTT 420
 GTGCGATATT TTCAATATCT CACTCCAGGT GTCCAGATCT ACGTTCTCCT CTCCATCTGC 480
 10 ATAGACCGGT TCTACACCAT CGTCTATCCT CTGAGCTTCA AGGTGTCCAG AGAAAAAGCC 540
 AAGAAAATGA TTGCGGCATC GTGGATCTTT GATGCAGGCT TTGTGACCCC TGTGCTCTTT 600
 TTCTATGGCT CCAACTGGGA CAGTCATTGT AACTATTTCC TCCCCTCCTC TTGGGAAGGC 660
 ACTGCCTACA CTGTCATCCA CTTCTTGGTG GGCTTTGTGA TTCCATCTGT CCTCATAATT 720
 TTATTTTACC AAAAGGTCAT AAAATATATT TGGAGAATAG GCACAGATGG CCGAACGGTG 780
 15 AGGAGGACAA TGAACATTGT CCCTCGGACA AAAGTGAAAA CTATCAAGAT GTTCCTCATT 840
 TTAAATCTGT TGTTTTTGCT CTCCTGGCTG CCTTTTCATG TAGCTCAGCT ATGGCACCCC 900
 CATGAACAAG ACTATAAGAA AAGTTCCCTT GTTTTCACAG CTATCACATG GATATCCTTT 960
 AGTTCTTCAG CCTCTAAACC TACTCTGTAT TCAATTTATA ATGCCAATTT TCGGAGAGGG 1020
 ATGAAAGAGA CTTTTTGCAT GTCCTCTATG AAATGTTACC GAAGCAATGC CTATACTATC 1080
 20 ACAACAAGTT CAAGGATGGC CAAAAAAAC TACGTTGGCA TTTCAGAAAT CCCTTCCATG 1140
 GCCAAACTA TTACCAAAGA CTCGATCTAT GACTCATTG ACAGAGAAGC CAAGGAAAAA 1200
 AAGCTTGCTT GGCCCATTA CTCAAATCCA CCAAATACTT TTGTCTAA 1248

(97) INFORMATION FOR SEQ ID NO:96:

- 25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 415 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

- 30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:

75

	Met	Val	Phe	Ala	His	Arg	Met	Asp	Asn	Ser	Lys	Pro	His	Leu	Ile	Ile	
	1				5					10					15		
	Pro	Thr	Leu	Leu	Val	Pro	Leu	Gln	Asn	Arg	Ser	Cys	Thr	Glu	Thr	Ala	
				20					25					30			
5	Thr	Pro	Leu	Pro	Ser	Gln	Tyr	Leu	Met	Glu	Leu	Ser	Glu	Glu	His	Ser	
			35					40					45				
	Trp	Met	Ser	Asn	Gln	Thr	Asp	Leu	His	Tyr	Val	Leu	Lys	Pro	Gly	Glu	
		50					55					60					
	Val	Ala	Thr	Ala	Ser	Ile	Phe	Phe	Gly	Ile	Leu	Trp	Leu	Phe	Ser	Ile	
10	65					70					75					80	
	Phe	Gly	Asn	Ser	Leu	Val	Cys	Leu	Val	Ile	His	Arg	Ser	Arg	Arg	Thr	
					85					90					95		
	Gln	Ser	Thr	Thr	Asn	Tyr	Phe	Val	Val	Ser	Met	Ala	Cys	Ala	Asp	Leu	
				100					105					110			
15	Leu	Ile	Ser	Val	Ala	Ser	Thr	Pro	Phe	Val	Leu	Leu	Gln	Phe	Thr	Thr	
			115					120					125				
	Gly	Arg	Trp	Thr	Leu	Gly	Ser	Ala	Thr	Cys	Lys	Val	Val	Arg	Tyr	Phe	
		130					135					140					
	Gln	Tyr	Leu	Thr	Pro	Gly	Val	Gln	Ile	Tyr	Val	Leu	Leu	Ser	Ile	Cys	
20	145					150					155					160	
	Ile	Asp	Arg	Phe	Tyr	Thr	Ile	Val	Tyr	Pro	Leu	Ser	Phe	Lys	Val	Ser	
					165					170					175		
	Arg	Glu	Lys	Ala	Lys	Lys	Met	Ile	Ala	Ala	Ser	Trp	Ile	Phe	Asp	Ala	
				180					185					190			
25	Gly	Phe	Val	Thr	Pro	Val	Leu	Phe	Phe	Tyr	Gly	Ser	Asn	Trp	Asp	Ser	
			195					200					205				
	His	Cys	Asn	Tyr	Phe	Leu	Pro	Ser	Ser	Trp	Glu	Gly	Thr	Ala	Tyr	Thr	
		210					215					220					
	Val	Ile	His	Phe	Leu	Val	Gly	Phe	Val	Ile	Pro	Ser	Val	Leu	Ile	Ile	
30	225					230					235					240	
	Leu	Phe	Tyr	Gln	Lys	Val	Ile	Lys	Tyr	Ile	Trp	Arg	Ile	Gly	Thr	Asp	
					245					250					255		
	Gly	Arg	Thr	Val	Arg	Arg	Thr	Met	Asn	Ile	Val	Pro	Arg	Thr	Lys	Val	
				260					265					270			
35	Lys	Thr	Ile	Lys	Met	Phe	Leu	Ile	Leu	Asn	Leu	Leu	Phe	Leu	Leu	Ser	
			275					280					285				

Trp Leu Pro Phe His Val Ala Gln Leu Trp His Pro His Glu Gln Asp
 290 295 300

Tyr Lys Lys Ser Ser Leu Val Phe Thr Ala Ile Thr Trp Ile Ser Phe
 305 310 315 320

5 Ser Ser Ser Ala Ser Lys Pro Thr Leu Tyr Ser Ile Tyr Asn Ala Asn
 325 330 335

Phe Arg Arg Gly Met Lys Glu Thr Phe Cys Met Ser Ser Met Lys Cys
 340 345 350

10 Tyr Arg Ser Asn Ala Tyr Thr Ile Thr Thr Ser Ser Arg Met Ala Lys
 355 360 365

Lys Asn Tyr Val Gly Ile Ser Glu Ile Pro Ser Met Ala Lys Thr Ile
 370 375 380

Thr Lys Asp Ser Ile Tyr Asp Ser Phe Asp Arg Glu Ala Lys Glu Lys
 385 390 395 400

15 Lys Leu Ala Trp Pro Ile Asn Ser Asn Pro Pro Asn Thr Phe Val
 405 410 415

(98) INFORMATION FOR SEQ ID NO:97:

- 20 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:97:

25 GGAAAGCTTA ACGATCCCCA GGAGCAACAT 30

(99) INFORMATION FOR SEQ ID NO:98:

- 30 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 31 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:98:

CTGGGATCCT ACGAGAGCAT TTTTCACACA G 31

35 (100) INFORMATION FOR SEQ ID NO:99:

- (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1842 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:99:

	ATGGGGCCCA	CCCTAGCGGT	TCCCACCCCC	TATGGCTGTA	TTGGCTGTAA	GCTACCCCAG	60
	CCAGAATACC	CACCGGCTCT	AATCATCTTT	ATGTTCTGCG	CGATGGTTAT	CACCATCGTT	120
	GTAGACCTAA	TCGGCAACTC	CATGGTCATT	TTGGCTGTGA	CGAAGAACAA	GAAGCTCCGG	180
10	AATTCTGGCA	ACATCTTCGT	GGTCAGTCTC	TCTGTGGCCG	ATATGCTGGT	GGCCATCTAC	240
	CCATACCCTT	TGATGCTGCA	TGCCATGTCC	ATTGGGGGCT	GGGATCTGAG	CCAGTTACAG	300
	TGCCAGATGG	TCGGGTTTCT	CACAGGGCTG	AGTGTGGTCG	GCTCCATCTT	CAACATCGTG	360
	GCAATCGCTA	TCAACCGTTA	CTGCTACATC	TGCCACAGCC	TCCAGTACGA	ACGGATCTTC	420
	AGTGTGCGCA	ATACCTGCAT	CTACCTGGTC	ATCACCTGGA	TCATGACCGT	CCTGGCTGTC	480
15	CTGCCCAACA	TGTACATTGG	CACCATCGAG	TACGATCCTC	GCACCTACAC	CTGCATCTTC	540
	AACTATCTGA	ACAACCCTGT	CTTCACTGTT	ACCATCGTCT	GCATCCACTT	CGTCCTCCCT	600
	CTCCTCATCG	TGGGTTTCTG	CTACGTGAGG	ATCTGGACCA	AAGTGCTGGC	GGCCCGTGAC	660
	CCTGCAGGGC	AGAATCCTGA	CAACCAACTT	GCTGAGGTTC	GCAATTTTCT	AACCATGTTT	720
	GTGATCTTCC	TCCTCTTTGC	AGTGTGCTGG	TGCCCTATCA	ACGTGCTCAC	TGTCTTGGTG	780
20	GCTGTCAGTC	CGAAGGAGAT	GGCAGGCAAG	ATCCCCAACT	GGCTTTATCT	TGCAGCCTAC	840
	TTCATAGCCT	ACTTCAACAG	CTGCCTCAAC	GCTGTGATCT	ACGGGCTCCT	CAATGAGAAT	900
	TTCCGAAGAG	AATACTGGAC	CATCTTCCAT	GCTATGCGGC	ACCCTATCAT	ATTCTTCCCT	960
	GGCCTCATCA	GTGATATTCG	TGAGATGCAG	GAGGCCCGTA	CCCTGGCCCCG	CGCCCGTGCC	1020
	CATGCTCGCG	ACCAAGCTCG	TGAACAAGAC	CGTGCCCATG	CCTGTCCTGC	TGTGGAGGAA	1080
25	ACCCCGATGA	ATGTCCGGAA	TGTTCCATTA	CCTGGTGATG	CTGCAGCTGG	CCACCCCGAC	1140
	CGTGCCTCTG	GCCACCCTAA	GCCCCATTCC	AGATCCTCCT	CTGCCTATCG	CAAATCTGCC	1200
	TCTACCCACC	ACAAGTCTGT	CTTTAGCCAC	TCCAAGGCTG	CCTCTGGTCA	CCTCAAGCCT	1260
	GTCTCTGGCC	ACTCCAAGCC	TGCCTCTGGT	CACCCCAAGT	CTGCCACTGT	CTACCCTAAG	1320
	CCTGCCTCTG	TCCATTTCAA	GGGTGACTCT	GTCCATTTCA	AGGGTGACTC	TGTCCATTTT	1380

AAGCCTGACT CTGTTTCATTT CAAGCCTGCT TCCAGCAACC CCAAGCCCAT CACTGGCCAC 1440
 CATGTCTCTG CTGGCAGCCA CTCCAAGTCT GCCTTCAGTG CTGCCACCAG CCACCCTAAA 1500
 CCCATCAAGC CAGCTACCAG CCATGCTGAG CCCACCACTG CTGACTATCC CAAGCCTGCC 1560
 ACTACCAGCC ACCCTAAGCC CGCTGCTGCT GACAACCCTG AGCTCTCTGC CTCCCATTGC 1620
 5 CCCGAGATCC CTGCCATTGC CCACCCTGTG TCTGACGACA GTGACCTCCC TGAGTCGGCC 1680
 TCTAGCCCTG CCGCTGGGCC CACCAAGCCT GCTGCCAGCC AGCTGGAGTC TGACACCATC 1740
 GCTGACCTTC CTGACCCTAC TGTAGTCACT ACCAGTACCA ATGATTACCA TGATGTCGTG 1800
 GTTGTGTGATG TTGAAGATGA TCCTGATGAA ATGGCTGTGT GA 1842

(101) INFORMATION FOR SEQ ID NO:100:

- 10 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 613 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant

- 15 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:

Met	Gly	Pro	Thr	Leu	Ala	Val	Pro	Thr	Pro	Tyr	Gly	Cys	Ile	Gly	Cys	1	5	10	15
Lys	Leu	Pro	Gln	Pro	Glu	Tyr	Pro	Pro	Ala	Leu	Ile	Ile	Phe	Met	Phe	20	25	30	
Cys	Ala	Met	Val	Ile	Thr	Ile	Val	Val	Asp	Leu	Ile	Gly	Asn	Ser	Met	35	40	45	
Val	Ile	Leu	Ala	Val	Thr	Lys	Asn	Lys	Lys	Leu	Arg	Asn	Ser	Gly	Asn	50	55	60	
Ile	Phe	Val	Val	Ser	Leu	Ser	Val	Ala	Asp	Met	Leu	Val	Ala	Ile	Tyr	65	70	75	80
Pro	Tyr	Pro	Leu	Met	Leu	His	Ala	Met	Ser	Ile	Gly	Gly	Trp	Asp	Leu	85	90	95	
Ser	Gln	Leu	Gln	Cys	Gln	Met	Val	Gly	Phe	Ile	Thr	Gly	Leu	Ser	Val	100	105	110	
Val	Gly	Ser	Ile	Phe	Asn	Ile	Val	Ala	Ile	Ala	Ile	Asn	Arg	Tyr	Cys	115	120	125	
Tyr	Ile	Cys	His	Ser	Leu	Gln	Tyr	Glu	Arg	Ile	Phe	Ser	Val	Arg	Asn	130	135	140	

Thr Cys Ile Tyr Leu Val Ile Thr Trp Ile Met Thr Val Leu Ala Val
 145 150 155 160
 Leu Pro Asn Met Tyr Ile Gly Thr Ile Glu Tyr Asp Pro Arg Thr Tyr
 165 170 175
 5 Thr Cys Ile Phe Asn Tyr Leu Asn Asn Pro Val Phe Thr Val Thr Ile
 180 185 190
 Val Cys Ile His Phe Val Leu Pro Leu Leu Ile Val Gly Phe Cys Tyr
 195 200 205
 10 Val Arg Ile Trp Thr Lys Val Leu Ala Ala Arg Asp Pro Ala Gly Gln
 210 215 220
 Asn Pro Asp Asn Gln Leu Ala Glu Val Arg Asn Phe Leu Thr Met Phe
 225 230 235 240
 Val Ile Phe Leu Leu Phe Ala Val Cys Trp Cys Pro Ile Asn Val Leu
 245 250 255
 15 Thr Val Leu Val Ala Val Ser Pro Lys Glu Met Ala Gly Lys Ile Pro
 260 265 270
 Asn Trp Leu Tyr Leu Ala Ala Tyr Phe Ile Ala Tyr Phe Asn Ser Cys
 275 280 285
 20 Leu Asn Ala Val Ile Tyr Gly Leu Leu Asn Glu Asn Phe Arg Arg Glu
 290 295 300
 Tyr Trp Thr Ile Phe His Ala Met Arg His Pro Ile Ile Phe Phe Pro
 305 310 315 320
 Gly Leu Ile Ser Asp Ile Arg Glu Met Gln Glu Ala Arg Thr Leu Ala
 325 330 335
 25 Arg Ala Arg Ala His Ala Arg Asp Gln Ala Arg Glu Gln Asp Arg Ala
 340 345 350
 His Ala Cys Pro Ala Val Glu Glu Thr Pro Met Asn Val Arg Asn Val
 355 360 365
 30 Pro Leu Pro Gly Asp Ala Ala Ala Gly His Pro Asp Arg Ala Ser Gly
 370 375 380
 His Pro Lys Pro His Ser Arg Ser Ser Ser Ala Tyr Arg Lys Ser Ala
 385 390 395 400
 Ser Thr His His Lys Ser Val Phe Ser His Ser Lys Ala Ala Ser Gly
 405 410 415
 35 His Leu Lys Pro Val Ser Gly His Ser Lys Pro Ala Ser Gly His Pro
 420 425 430
 Lys Ser Ala Thr Val Tyr Pro Lys Pro Ala Ser Val His Phe Lys Gly

	435	440	445
	Asp Ser Val His Phe Lys Gly Asp Ser Val His Phe Lys Pro Asp Ser		
	450	455	460
5	Val His Phe Lys Pro Ala Ser Ser Asn Pro Lys Pro Ile Thr Gly His		
	465	470	475 480
	His Val Ser Ala Gly Ser His Ser Lys Ser Ala Phe Ser Ala Ala Thr		
		485	490 495
	Ser His Pro Lys Pro Ile Lys Pro Ala Thr Ser His Ala Glu Pro Thr		
		500	505 510
10	Thr Ala Asp Tyr Pro Lys Pro Ala Thr Thr Ser His Pro Lys Pro Ala		
		515	520 525
	Ala Ala Asp Asn Pro Glu Leu Ser Ala Ser His Cys Pro Glu Ile Pro		
		530	535 540
15	Ala Ile Ala His Pro Val Ser Asp Asp Ser Asp Leu Pro Glu Ser Ala		
		545	550 555 560
	Ser Ser Pro Ala Ala Gly Pro Thr Lys Pro Ala Ala Ser Gln Leu Glu		
		565	570 575
	Ser Asp Thr Ile Ala Asp Leu Pro Asp Pro Thr Val Val Thr Thr Ser		
		580	585 590
20	Thr Asn Asp Tyr His Asp Val Val Val Val Asp Val Glu Asp Asp Pro		
		595	600 605
	Asp Glu Met Ala Val		
		610	

(102) INFORMATION FOR SEQ ID NO:101:

- 25 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 32 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- 30 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:101:

TCCAAGCTTC GCCATGGGAC ATAACGGGAG CT

32

(103) INFORMATION FOR SEQ ID NO:102:

- 35 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:102:

CGTGAATTCC AAGAATTTAC AATCCTTGCT

30

5 (104) INFORMATION FOR SEQ ID NO:103:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1548 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:103:

ATGGGACATA	ACGGGAGCTG	GATCTCTCCA	AATGCCAGCG	AGCCGCACAA	CGCGTCCGGC	60
GCCGAGGCTG	CGGGTGTGAA	CCGCAGCGCG	CTCGGGGAGT	TCGGCGAGGC	GCAGCTGTAC	120
15 CGCCAGTTCA	CCACCACCGT	GCAGGTCGTC	ATCTTCATAG	GCTCGCTGCT	CGGAAACTTC	180
ATGGTGTAT	GGTCAACTTG	CCGCACAACC	GTGTTCAAAT	CTGTCACCAA	CAGGTTTCATT	240
AAAAACCTGG	CCTGCTCGGG	GATTTGTGCC	AGCCTGGTCT	GTGTGCCCTT	CGACATCATC	300
CTCAGCACCA	GTCCTCACTG	TTGCTGGTGG	ATCTACACCA	TGCTCTTCTG	CAAGGTCGTC	360
AAATTTTTC	ACAAAGTATT	CTGCTCTGTG	ACCATCCTCA	GCTTCCCTGC	TATTGCTTTG	420
20 GACAGGTACT	ACTCAGTCCT	CTATCCACTG	GAGAGGAAAA	TATCTGATGC	CAAGTCCCGT	480
GAACTGGTGA	TGTACATCTG	GGCCCATGCA	GTGGTGGCCA	GTGTCCCTGT	GTTTGCAGTA	540
ACCAATGTGG	CTGACATCTA	TGCCACGTCC	ACCTGCACGG	AAGTCTGGAG	CAACTCCTTG	600
GGCCACCTGG	TGTACGTTCT	GGTGTATAAC	ATCACCACGG	TCATTGTGCC	TGTGGTGGTG	660
GTGTTCTCT	TCTTGATACT	GATCCGACGG	GCCCTGAGTG	CCAGCCAGAA	GAAGAAGGTC	720
25 ATCATAGCAG	CGCTCCGGAC	CCCACAGAAC	ACCATCTCTA	TTCCCTATGC	CTCCCAGCGG	780
GAGGCCGAGC	TGCACGCCAC	CCTGCTCTCC	ATGGTGATGG	TCTTCATCTT	GTGTAGCGTG	840
CCCTATGCCA	CCCTGGTCGT	CTACCAGACT	GTGCTCAATG	TCCCTGACAC	TTCCGTCTTC	900
TTGCTGCTCA	CTGCTGTTTG	GCTGCCCAAA	GTCTCCCTGC	TGGCAAACCC	TGTTCTCTTT	960
CTTACTGTGA	ACAAATCTGT	CCGCAAGTGC	TTGATAGGGA	CCCTGGTGCA	ACTACACCAC	1020
30 CGGTACAGTC	GCCGTAATGT	GGTCAGTACA	GGGAGTGGCA	TGGCTGAGGC	CAGCCTGGAA	1080

CCCAGCATAC GCTCGGGTAG CCAGCTCCTG GAGATGTTCC ACATTGGGCA GCAGCAGATC 1140
 TTTAAGCCCCA CAGAGGATGA GGAAGAGAGT GAGGCCAAGT ACATTGGCTC AGCTGACTTC 1200
 CAGGCCAAGG AGATATTTAG CACCTGCCTC GAGGGAGAGC AGGGGCCACA GTTTGCGCCC 1260
 TCTGCCCCAC CCCTGAGCAC AGTGGACTCT GTATCCCAGG TGGCACCGGC AGCCCCTGTG 1320
 5 GAACCTGAAA CATTCCTGA TAAGTATTC CTGCAGTTTG GCTTTGGGCC TTTTGAGTTG 1380
 CCTCCTCAGT GGCTCTCAGA GACCCGAAAC AGCAAGAAGC GGCTGCTTCC CCCCTTGGGC 1440
 AACACCCAG AAGAGCTGAT CCAGACAAAG GTGCCCAAGG TAGGCAGGGT GGAGCGGAAG 1500
 ATGAGCAGAA ACAATAAAGT GAGCATTTTT CCAAAGGTGG ATTCCTAG 1548

(105) INFORMATION FOR SEQ ID NO:104:

10 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 515 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

15 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:104:

	Met	Gly	His	Asn	Gly	Ser	Trp	Ile	Ser	Pro	Asn	Ala	Ser	Glu	Pro	His	
	1				5					10					15		
	Asn	Ala	Ser	Gly	Ala	Glu	Ala	Ala	Gly	Val	Asn	Arg	Ser	Ala	Leu	Gly	
20				20					25					30			
	Glu	Phe	Gly	Glu	Ala	Gln	Leu	Tyr	Arg	Gln	Phe	Thr	Thr	Thr	Val	Gln	
			35					40					45				
	Val	Val	Ile	Phe	Ile	Gly	Ser	Leu	Leu	Gly	Asn	Phe	Met	Val	Leu	Trp	
			50				55					60					
25	Ser	Thr	Cys	Arg	Thr	Thr	Val	Phe	Lys	Ser	Val	Thr	Asn	Arg	Phe	Ile	
	65					70					75				80		
	Lys	Asn	Leu	Ala	Cys	Ser	Gly	Ile	Cys	Ala	Ser	Leu	Val	Cys	Val	Pro	
				85					90					95			
	Phe	Asp	Ile	Ile	Leu	Ser	Thr	Ser	Pro	His	Cys	Cys	Trp	Trp	Ile	Tyr	
30				100					105					110			
	Thr	Met	Leu	Phe	Cys	Lys	Val	Val	Lys	Phe	Leu	His	Lys	Val	Phe	Cys	
			115					120					125				
	Ser	Val	Thr	Ile	Leu	Ser	Phe	Pro	Ala	Ile	Ala	Leu	Asp	Arg	Tyr	Tyr	
			130				135					140					

83

	Ser Val Leu Tyr Pro Leu Glu Arg Lys Ile Ser Asp Ala Lys Ser Arg	145	150	155	160
	Glu Leu Val Met Tyr Ile Trp Ala His Ala Val Val Ala Ser Val Pro	165	170	175	
5	Val Phe Ala Val Thr Asn Val Ala Asp Ile Tyr Ala Thr Ser Thr Cys	180	185	190	
	Thr Glu Val Trp Ser Asn Ser Leu Gly His Leu Val Tyr Val Leu Val	195	200	205	
10	Tyr Asn Ile Thr Thr Val Ile Val Pro Val Val Val Val Phe Leu Phe	210	215	220	
	Leu Ile Leu Ile Arg Arg Ala Leu Ser Ala Ser Gln Lys Lys Lys Val	225	230	235	240
	Ile Ile Ala Ala Leu Arg Thr Pro Gln Asn Thr Ile Ser Ile Pro Tyr	245	250	255	
15	Ala Ser Gln Arg Glu Ala Glu Leu His Ala Thr Leu Leu Ser Met Val	260	265	270	
	Met Val Phe Ile Leu Cys Ser Val Pro Tyr Ala Thr Leu Val Val Tyr	275	280	285	
20	Gln Thr Val Leu Asn Val Pro Asp Thr Ser Val Phe Leu Leu Leu Thr	290	295	300	
	Ala Val Trp Leu Pro Lys Val Ser Leu Leu Ala Asn Pro Val Leu Phe	305	310	315	320
	Leu Thr Val Asn Lys Ser Val Arg Lys Cys Leu Ile Gly Thr Leu Val	325	330	335	
25	Gln Leu His His Arg Tyr Ser Arg Arg Asn Val Val Ser Thr Gly Ser	340	345	350	
	Gly Met Ala Glu Ala Ser Leu Glu Pro Ser Ile Arg Ser Gly Ser Gln	355	360	365	
30	Leu Leu Glu Met Phe His Ile Gly Gln Gln Gln Ile Phe Lys Pro Thr	370	375	380	
	Glu Asp Glu Glu Glu Ser Glu Ala Lys Tyr Ile Gly Ser Ala Asp Phe	385	390	395	400
	Gln Ala Lys Glu Ile Phe Ser Thr Cys Leu Glu Gly Glu Gln Gly Pro	405	410	415	
35	Gln Phe Ala Pro Ser Ala Pro Pro Leu Ser Thr Val Asp Ser Val Ser	420	425	430	
	Gln Val Ala Pro Ala Ala Pro Val Glu Pro Glu Thr Phe Pro Asp Lys				

435

440

445

Tyr Ser Leu Gln Phe Gly Phe Gly Pro Phe Glu Leu Pro Pro Gln Trp
 450 455 460

5 Leu Ser Glu Thr Arg Asn Ser Lys Lys Arg Leu Leu Pro Pro Leu Gly
 465 470 475 480

Asn Thr Pro Glu Glu Leu Ile Gln Thr Lys Val Pro Lys Val Gly Arg
 485 490 495

Val Glu Arg Lys Met Ser Arg Asn Asn Lys Val Ser Ile Phe Pro Lys
 500 505 510

10 Val Asp Ser
 515

(106) INFORMATION FOR SEQ ID NO:105:

15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 29 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:105:

20 GGAGAATTCA CTAGGCGAGG CGCTCCATC

29

(107) INFORMATION FOR SEQ ID NO:106:

25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:106:

GGAGGATCCA GGAAACCTTA GGCCGAGTCC

30

30 (108) INFORMATION FOR SEQ ID NO:107:

35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1164 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:107:

ATGAATCGGC ACCATCTGCA GGATCACTTT CTGGAAATAG ACAAGAAGAA CTGCTGTGTG 60
 TTCCGAGATG ACTTCATTGC CAAGGTGTTG CCGCCGGTGT TGGGGCTGGA GTTTATCTTT 120
 GGGCTTCTGG GCAATGGCCT TGCCCTGTGG ATTTTCTGTT TCCACCTCAA GTCCTGGAAA 180
 5 TCCAGCCGGA TTTTCCTGTT CAACCTGGCA GTAGCTGACT TTCTACTGAT CATCTGCCTG 240
 CCGTTCGTGA TGGACTACTA TGTGCGGCGT TCAGACTGGA ACTTTGGGGA CATCCCTTGC 300
 CGGCTGGTGC TCTTCATGTT TGCCATGAAC CGCCAGGGCA GCATCATCTT CCTCACGGTG 360
 GTGGCGGTAG ACAGGTATTT CCGGGTGGTC CATCCCCACC ACGCCCTGAA CAAGATCTCC 420
 AATTGGACAG CAGCCATCAT CTCTTGCTT CTGTGGGGCA TCACTGTTGG CCTAACAGTC 480
 10 CACCTCCTGA AGAAGAAGTT GCTGATCCAG AATGGCCCTG CAAATGTGTG CATCAGCTTC 540
 AGCATCTGCC ATACCTTCCG GTGGCACGAA GCTATGTTCC TCCTGGAGTT CCTCCTGCCC 600
 CTGGGCATCA TCCTGTTCTG CTCAGCCAGA ATTATCTGGA GCCTGCGGCA GAGACAAATG 660
 GACCGGCATG CCAAGATCAA GAGAGCCATC ACCTTCATCA TGGTGGTGGC CATCGTCTTT 720
 GTCATCTGCT TCCTTCCCAG CGTGTTGTG CGGATCCGCA TCTTCTGGCT CCTGCACACT 780
 15 TCGGGCACGC AGAATTGTGA AGTGATCCGC TCGGTGGACC TGGCGTTCTT TATCACTCTC 840
 AGCTTCACCT ACATGAACAG CATGCTGGAC CCCGTGGTGT ACTACTTCTC CAGCCCATCC 900
 TTTCCCAACT TCTTCTCCAC TTTGATCAAC CGCTGCCTCC AGAGGAAGAT GACAGGTGAG 960
 CCAGATAATA ACCGCAGCAC GAGCGTCGAG CTCACAGGGG ACCCCAACAA AACCAGAGGC 1020
 GCTCCAGAGG CGTTAATGGC CAACTCCGGT GAGCCATGGA GCCCCTCTTA TCTGGGCCCA 1080
 20 ACCTCAAATA ACCATTCCAA GAAGGGACAT TGTCACCAAG AACCAGCATC TCTGGAGAAA 1140
 CAGTTGGGCT GTTGCATCGA GTAA 1164

(109) INFORMATION FOR SEQ ID NO:108:

(i) SEQUENCE CHARACTERISTICS:

25

- (A) LENGTH: 387 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:108:

30

Met Asn Arg His His Leu Gln Asp His Phe Leu Glu Ile Asp Lys Lys

1	5	10	15
Asn Cys Cys Val Phe Arg Asp Asp Phe Ile Ala Lys Val Leu Pro Pro	20	25	30
Val Leu Gly Leu Glu Phe Ile Phe Gly Leu Leu Gly Asn Gly Leu Ala	35	40	45
Leu Trp Ile Phe Cys Phe His Leu Lys Ser Trp Lys Ser Ser Arg Ile	50	55	60
Phe Leu Phe Asn Leu Ala Val Ala Asp Phe Leu Leu Ile Ile Cys Leu	65	70	75
Pro Phe Val Met Asp Tyr Tyr Val Arg Arg Ser Asp Trp Asn Phe Gly	85	90	95
Asp Ile Pro Cys Arg Leu Val Leu Phe Met Phe Ala Met Asn Arg Gln	100	105	110
Gly Ser Ile Ile Phe Leu Thr Val Val Ala Val Asp Arg Tyr Phe Arg	115	120	125
Val Val His Pro His His Ala Leu Asn Lys Ile Ser Asn Trp Thr Ala	130	135	140
Ala Ile Ile Ser Cys Leu Leu Trp Gly Ile Thr Val Gly Leu Thr Val	145	150	155
His Leu Leu Lys Lys Lys Leu Leu Ile Gln Asn Gly Pro Ala Asn Val	165	170	175
Cys Ile Ser Phe Ser Ile Cys His Thr Phe Arg Trp His Glu Ala Met	180	185	190
Phe Leu Leu Glu Phe Leu Leu Pro Leu Gly Ile Ile Leu Phe Cys Ser	195	200	205
Ala Arg Ile Ile Trp Ser Leu Arg Gln Arg Gln Met Asp Arg His Ala	210	215	220
Lys Ile Lys Arg Ala Ile Thr Phe Ile Met Val Val Ala Ile Val Phe	225	230	235
Val Ile Cys Phe Leu Pro Ser Val Val Val Arg Ile Arg Ile Phe Trp	245	250	255
Leu Leu His Thr Ser Gly Thr Gln Asn Cys Glu Val Tyr Arg Ser Val	260	265	270
Asp Leu Ala Phe Phe Ile Thr Leu Ser Phe Thr Tyr Met Asn Ser Met	275	280	285
Leu Asp Pro Val Val Tyr Tyr Phe Ser Ser Pro Ser Phe Pro Asn Phe	290	295	300

87

Phe Ser Thr Leu Ile Asn Arg Cys Leu Gln Arg Lys Met Thr Gly Glu
 305 310 315 320
 Pro Asp Asn Asn Arg Ser Thr Ser Val Glu Leu Thr Gly Asp Pro Asn
 325 330 335
 5 Lys Thr Arg Gly Ala Pro Glu Ala Leu Met Ala Asn Ser Gly Glu Pro
 340 345 350
 Trp Ser Pro Ser Tyr Leu Gly Pro Thr Ser Asn Asn His Ser Lys Lys
 355 360 365
 10 Gly His Cys His Gln Glu Pro Ala Ser Leu Glu Lys Gln Leu Gly Cys
 370 375 380
 Cys Ile Glu
 385

(110) INFORMATION FOR SEQ ID NO:109:

- 15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 37 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: DNA (genomic)
 20 (iv) ANTI-SENSE: NO
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:109:

ACCATGGCTT GCAATGGCAG TCGGCCAGG GGGCACT

37

(111) INFORMATION FOR SEQ ID NO:110:

- 25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 39 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: DNA (genomic)
 30 (iv) ANTI-SENSE: YES
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:110:

CGACCAGGAC AAACAGCATC TTGGTCACTT GTCTCCGGC

39

(112) INFORMATION FOR SEQ ID NO:111:

- 35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 39 base pairs
 (B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

5 (iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:111:

GACCAAGATG CTGTTTGTCC TGGTCGTGGT GTTTGGCAT

39

(113) INFORMATION FOR SEQ ID NO:112:

10 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

15 (iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:112:

CGGAATTCAG GATGGATCGG TCTCTTGCTG CGCCT

35

(114) INFORMATION FOR SEQ ID NO:113:

20 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1212 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:113:

ATGGCTTGCA ATGGCAGTGC GGCCAGGGGG CACTTTGACC CTGAGGACTT GAACCTGACT 60

GACGAGGCAC TGAGACTCAA GTACCTGGGG CCCAGCAGA CAGAGCTGTT CATGCCCATC 120

TGTGCCACAT ACCTGCTGAT CTCGTGGTG GCGCTGTGG GCAATGGGCT GACCTGTCTG 180

GTCATCCTGC GCCACAAGGC CATGCGCAGC CCTACCAACT ACTACCTCTT CAGCCTGGCC 240

30 GTGTCGGACC TGCTGGTGCT GCTGGTGGGC CTGCCCTGG AGCTCTATGA GATGTGGCAC 300

AACTACCCCT TCCTGCTGGG CGTTGGTGGC TGCTATTTCC GCACGCTACT GTTTGAGATG 360

GTCTGCCTGG CCTCAGTGCT CAACGTCAGT GCCCTGAGCG TGAACGCTA TGTGGCCGTG 420

GTGCACCCAC TCCAGGCCAG GTCCATGGTG ACGCGGGCCC ATGTGCGCCG AGTGCTTGGG 480

GCCGTCTGGG GTCTTGCCAT GCTCTGCTCC CTGCCCAACA CCAGCCTGCA CGGCATCCGG 540
 CAGCTGCACG TGCCCTGCCG GGGCCCAGTG CCAGACTCAG CTGTTTGCAT GCTGGTCCGC 600
 CCACGGGCCC TCTACAACAT GGTAGTGCAG ACCACCGCGC TGCTCTTCTT CTGCCTGCCC 660
 ATGGCCATCA TGAGCGTGCT CTACCTGCTC ATTGGGCTGC GACTGCGGCG GGAGAGGCTG 720
 5 CTGCTCATGC AGGAGGCCAA GGGCAGGGG TCTGCAGCAG CCAGGTCCAG ATACACCTGC 780
 AGGCTCCAGC AGCACGATCG GGGCCGGAGA CAAGTGACCA AGATGCTGTT TGTCTGGTTC 840
 GTGGTGTGTTG GCATCTGCTG GGCCCCGTTT CACGCCGACC GCGTCATGTG GAGCGTCGTG 900
 TCACAGTGGA CAGATGGCCT GCACCTGGCC TTCCAGCACG TGCACGTCAT CTCCGGCATC 960
 TTCTTCTACC TGGGCTCGGC GGCCAACCCC GTGCTCTATA GCCTCATGTC CAGCCGCTTC 1020
 10 CGAGAGACCT TCCAGGAGGC CCTGTGCCTC GGGGCCTGCT GCCATCGCCT CAGACCCCGC 1080
 CACAGCTCCC ACAGCCTCAG CAGGATGACC ACAGGCAGCA CCCTGTGTGA TGTGGGCTCC 1140
 CTGGGCAGCT GGGTCCACCC CCTGGCTGGG AACGATGGCC CAGAGGCGCA GCAAGAGACC 1200
 GATCCATCCT GA 1212

(115) INFORMATION FOR SEQ ID NO:114:

- 15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 403 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

- 20 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:114:

Met Ala Cys Asn Gly Ser Ala Ala Arg Gly His Phe Asp Pro Glu Asp
 1 5 10 15
 Leu Asn Leu Thr Asp Glu Ala Leu Arg Leu Lys Tyr Leu Gly Pro Gln
 20 25 30
 Gln Thr Glu Leu Phe Met Pro Ile Cys Ala Thr Tyr Leu Leu Ile Phe
 35 40 45
 Val Val Gly Ala Val Gly Asn Gly Leu Thr Cys Leu Val Ile Leu Arg
 50 55 60
 30 His Lys Ala Met Arg Thr Pro Thr Asn Tyr Tyr Leu Phe Ser Leu Ala
 65 70 75 80
 Val Ser Asp Leu Leu Val Leu Leu Val Gly Leu Pro Leu Glu Leu Tyr
 85 90 95

	Glu Met Trp His Asn Tyr Pro Phe Leu Leu Gly Val Gly Gly Cys Tyr	
	100	105 110
	Phe Arg Thr Leu Leu Phe Glu Met Val Cys Leu Ala Ser Val Leu Asn	
	115	120 125
5	Val Thr Ala Leu Ser Val Glu Arg Tyr Val Ala Val Val His Pro Leu	
	130	135 140
	Gln Ala Arg Ser Met Val Thr Arg Ala His Val Arg Arg Val Leu Gly	
	145	150 155 160
10	Ala Val Trp Gly Leu Ala Met Leu Cys Ser Leu Pro Asn Thr Ser Leu	
	165	170 175
	His Gly Ile Arg Gln Leu His Val Pro Cys Arg Gly Pro Val Pro Asp	
	180	185 190
	Ser Ala Val Cys Met Leu Val Arg Pro Arg Ala Leu Tyr Asn Met Val	
	195	200 205
15	Val Gln Thr Thr Ala Leu Leu Phe Phe Cys Leu Pro Met Ala Ile Met	
	210	215 220
	Ser Val Leu Tyr Leu Leu Ile Gly Leu Arg Leu Arg Arg Glu Arg Leu	
	225	230 235 240
20	Leu Leu Met Gln Glu Ala Lys Gly Arg Gly Ser Ala Ala Ala Arg Ser	
	245	250 255
	Arg Tyr Thr Cys Arg Leu Gln Gln His Asp Arg Gly Arg Arg Gln Val	
	260	265 270
	Thr Lys Met Leu Phe Val Leu Val Val Val Phe Gly Ile Cys Trp Ala	
	275	280 285
25	Pro Phe His Ala Asp Arg Val Met Trp Ser Val Val Ser Gln Trp Thr	
	290	295 300
	Asp Gly Leu His Leu Ala Phe Gln His Val His Val Ile Ser Gly Ile	
	305	310 315 320
30	Phe Phe Tyr Leu Gly Ser Ala Ala Asn Pro Val Leu Tyr Ser Leu Met	
	325	330 335
	Ser Ser Arg Phe Arg Glu Thr Phe Gln Glu Ala Leu Cys Leu Gly Ala	
	340	345 350
	Cys Cys His Arg Leu Arg Pro Arg His Ser Ser His Ser Leu Ser Arg	
	355	360 365
35	Met Thr Thr Gly Ser Thr Leu Cys Asp Val Gly Ser Leu Gly Ser Trp	
	370	375 380
	Val His Pro Leu Ala Gly Asn Asp Gly Pro Glu Ala Gln Gln Glu Thr	

385

390

395

400

Asp Pro Ser

(116) INFORMATION FOR SEQ ID NO:115:

- 5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

- 10 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:115:

GGAAGCTTCA GGCCCAAAGA TGGGGAACAT

30

(117) INFORMATION FOR SEQ ID NO:116:

- 15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:116:

GTGGATCCAC CCGCGGAGGA CCCAGGCTAG

30

(118) INFORMATION FOR SEQ ID NO:117:

- 25 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1098 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:117:

- 30 ATGGGGAACA TCACTGCAGA CAACTCCTCG ATGAGCTGTA CCATCGACCA TACCATCCAC 60
CAGACGCTGG CCCCGGTGGT CTATGTTACC GTGCTGGTGG TGGGCTTCCC GGCCAAC TGC 120
CTGTCCCTCT ACTTCGGCTA CCTGCAGATC AAGGCCCGGA ACGAGCTGGG CGTGTACCTG 180
TGCAACCTGA CGGTGGCCGA CCTCTTCTAC ATCTGCTCGC TGCCCTTCTG GCTGCAGTAC 240
GTGCTGCAGC ACGACAACTG GTCTCACGGC GACCTGTCCT GCCAGGTGTG CGGCATCCTC 300
35 CTGTACGAGA ACATCTACAT CAGCGTGGGC TTCCTCTGCT GCATCTCCGT GGACCGCTAC 360

CTGGCTGTGG CCCATCCCTT CCGCTTCCAC CAGTTCGGA CCCTGAAGGC GGCCGTCGGC 420
 GTCAGCGTGG TCATCTGGGC CAAGGAGCTG CTGACCAGCA TCTACTTCCT GATGCACGAG 480
 GAGGTCATCG AGGACGAGAA CCAGCACCGC GTGTGCTTTG AGCACTACCC CATCCAGGCA 540
 TGGCAGCGCG CCATCAACTA CTACCGCTTC CTGGTGGGCT TCCTCTTCCC CATCTGCCTG 600
 5 CTGCTGGCGT CCTACCAGGG CATCCTGCGC GCCGTGCGCC GGAGCCACGG CACCCAGAAG 660
 AGCCGCAAGG ACCAGATCCA GCGGCTGGTG CTCAGCACCG TGGTCATCTT CCTGGCCTGC 720
 TTCCTGCCCT ACCACGTGTT GCTGCTGGTG CGCAGCGTCT GGGAGGCCAG CTGCGACTTC 780
 GCCAAGGGCG TTTTCAACGC CTACCACTTC TCCCTCCTGC TCACCAGCTT CAACTGCGTC 840
 GCCGACCCCG TGCTCTACTG CTTGTCAGC GAGACCACCC ACCGGGACCT GGCCCGCCTC 900
 10 CGCGGGGCCT GCCTGGCCTT CCTCACCTGC TCCAGGACCG GCCGGGCCAG GGAGGCCTAC 960
 CCCTGGGTG CCCCCGAGGC CTCCGGGAAA AGCGGGGCCC AGGGTGAGGA GCCCGAGCTG 1020
 TTGACCAAGC TCCACCCGGC CTTCCAGACC CTAACCTCGC CAGGGTCGGG CGGGTTCCCC 1080
 ACGGGCAGGT TGGCCTAG 1098

(119) INFORMATION FOR SEQ ID NO:118:

- 15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 365 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

- 20 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:118:

Met Gly Asn Ile Thr Ala Asp Asn Ser Ser Met Ser Cys Thr Ile Asp
 1 5 10 15

25 His Thr Ile His Gln Thr Leu Ala Pro Val Val Tyr Val Thr Val Leu
 20 25 30

Val Val Gly Phe Pro Ala Asn Cys Leu Ser Leu Tyr Phe Gly Tyr Leu
 35 40 45

Gln Ile Lys Ala Arg Asn Glu Leu Gly Val Tyr Leu Cys Asn Leu Thr
 50 55 60

30 Val Ala Asp Leu Phe Tyr Ile Cys Ser Leu Pro Phe Trp Leu Gln Tyr
 65 70 75 80

Val Leu Gln His Asp Asn Trp Ser His Gly Asp Leu Ser Cys Gln Val
 85 90 95

93

	Cys Gly Ile Leu Leu Tyr Glu Asn Ile Tyr Ile Ser Val Gly Phe Leu	
	100	105 110
	Cys Cys Ile Ser Val Asp Arg Tyr Leu Ala Val Ala His Pro Phe Arg	
	115	120 125
5	Phe His Gln Phe Arg Thr Leu Lys Ala Ala Val Gly Val Ser Val Val	
	130	135 140
	Ile Trp Ala Lys Glu Leu Leu Thr Ser Ile Tyr Phe Leu Met His Glu	
	145	150 155 160
10	Glu Val Ile Glu Asp Glu Asn Gln His Arg Val Cys Phe Glu His Tyr	
	165	170 175
	Pro Ile Gln Ala Trp Gln Arg Ala Ile Asn Tyr Tyr Arg Phe Leu Val	
	180	185 190
	Gly Phe Leu Phe Pro Ile Cys Leu Leu Leu Ala Ser Tyr Gln Gly Ile	
	195	200 205
15	Leu Arg Ala Val Arg Arg Ser His Gly Thr Gln Lys Ser Arg Lys Asp	
	210	215 220
	Gln Ile Gln Arg Leu Val Leu Ser Thr Val Val Ile Phe Leu Ala Cys	
	225	230 235 240
20	Phe Leu Pro Tyr His Val Leu Leu Leu Val Arg Ser Val Trp Glu Ala	
	245	250 255
	Ser Cys Asp Phe Ala Lys Gly Val Phe Asn Ala Tyr His Phe Ser Leu	
	260	265 270
	Leu Leu Thr Ser Phe Asn Cys Val Ala Asp Pro Val Leu Tyr Cys Phe	
	275	280 285
25	Val Ser Glu Thr Thr His Arg Asp Leu Ala Arg Leu Arg Gly Ala Cys	
	290	295 300
	Leu Ala Phe Leu Thr Cys Ser Arg Thr Gly Arg Ala Arg Glu Ala Tyr	
	305	310 315 320
30	Pro Leu Gly Ala Pro Glu Ala Ser Gly Lys Ser Gly Ala Gln Gly Glu	
	325	330 335
	Glu Pro Glu Leu Leu Thr Lys Leu His Pro Ala Phe Gln Thr Pro Asn	
	340	345 350
	Ser Pro Gly Ser Gly Gly Phe Pro Thr Gly Arg Leu Ala	
	355	360 365

35 (120) INFORMATION FOR SEQ ID NO:119:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 26 base pairs

(B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:119:

GACCTCGAGT CCTTCTACAC CTCATC

26

(121) INFORMATION FOR SEQ ID NO:120:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:120:

15 TGCTCTAGAT TCCAGATAGG TGAAAACCTTG

30

(122) INFORMATION FOR SEQ ID NO:121:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1416 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:121:

ATGGATATTC TTTGTGAAGA AAATACTTCT TTGAGCTCAA CTACGAACTC CCTAATGCAA 60

25 TTAAATGATG ACAACAGGCT CTACAGTAAT GACTTTAACT CCGGAGAAGC TAACACTTCT 120

GATGCATTTA ACTGGACAGT CGACTCTGAA AATCGAACCA ACCTTTCCTG TGAAGGGTGC 180

CTCTCACCGT CGTGTCTCTC CTTACTTCAT CTCCAGGAAA AAAACTGGTC TGCTTTACTG 240

ACAGCCGTAG TGATTATTCT AACTATTGCT GGAAACATAC TCGTCATCAT GGCAGTGTCC 300

CTAGAGAAAA AGCTGCAGAA TGCCACCAAC TATTTCTCTGA TGTCACCTGC CATAGCTGAT 360

30 ATGCTGCTGG GTTTCCTTGT CATGCCCGTG TCCATGTTAA CCATCCTGTA TGGGTACCGG 420

TGGCCTCTGC CGAGCAAGCT TTGTGCAGTC TGGATTTACC TGGACGTGCT CTTCTCCACG 480

GCCTCCATCA TGCACCTCTG CGCCATCTCG CTGGACCGCT ACGTCGCCAT CCAGAATCCC 540

ATCCACCACA GCCGCTTCAA CTCCAGAACT AAGGCATTTT TGAAAATCAT TGCTGTTTGG 600

ACCATATCAG TAGGTATATC CATGCCAATA CCAGTCTTTG GGCTACAGGA CGATTCGAAG 660
 GTCTTTAAGG AGGGGAGTTG CTTACTCGCC GATGATAACT TTGTCCTGAT CGGCTCTTTT 720
 GTGTCATTTT TCATTCCCTT AACCATCATG GTGATCACCT ACTTTCTAAC TATCAAGTCA 780
 CTCCAGAAAG AAGCTACTTT GTGTGTAAGT GATCTTGGCA CACGGGCCAA ATTAGCTTCT 840
 5 TTCAGCTTCC TCCCTCAGAG TTCTTTGTCT TCAGAAAAGC TCTTCAGCG GTCGATCCAT 900
 AGGGAGCCAG GGTCTACAC AGGCAGGAGG ACTATGCAGT CCATCAGCAA TGAGCAAAAG 960
 GCATGCAAGG TGCTGGGCAT CGTCTTCTTC CTGTTTGTGG TGATGTGGTG CCCTTTCTTC 1020
 ATCACAAACA TCATGGCCGT CATCTGCAA GAGTCCTGCA ATGAGGATGT CATTGGGGCC 1080
 CTGCTCAATG TGTGTGTTG GATCGTTAT CTCTCTTCAG CAGTCAACCC ACTAGTCTAC 1140
 10 ACACTGTTCA ACAAGACCTA TAGGTCAGCC TTTTCACGGT ATATTCAGTG TCAGTACAAG 1200
 GAAAACAAAA AACCATTGCA GTTAATTTTA GTGAACACAA TACCGGCTTT GGCCTACAAG 1260
 TCTAGCCAAC TTCAAATGGG AAAAAAAG AATTCAAAGC AAGATGCCAA GACAACAGAT 1320
 AATGACTGCT CAATGGTTGC TCTAGGAAAG CAGTATTCTG AAGAGGCTTC TAAAGACAAT 1380
 AGCGACGGAG TGAATGAAAA GGTGAGCTGT GTGTGA 1416

15 (123) INFORMATION FOR SEQ ID NO:122:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 471 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant

20

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:122:

Met Asp Ile Leu Cys Glu Glu Asn Thr Ser Leu Ser Ser Thr Thr Asn
 1 5 10 15
 25 Ser Leu Met Gln Leu Asn Asp Asp Asn Arg Leu Tyr Ser Asn Asp Phe
 20 25 30
 Asn Ser Gly Glu Ala Asn Thr Ser Asp Ala Phe Asn Trp Thr Val Asp
 35 40 45
 30 Ser Glu Asn Arg Thr Asn Leu Ser Cys Glu Gly Cys Leu Ser Pro Ser
 50 55 60
 Cys Leu Ser Leu Leu His Leu Gln Glu Lys Asn Trp Ser Ala Leu Leu
 65 70 75 80

	Thr	Ala	Val	Val	Ile	Ile	Leu	Thr	Ile	Ala	Gly	Asn	Ile	Leu	Val	Ile	
					85					90				95			
	Met	Ala	Val	Ser	Leu	Glu	Lys	Lys	Leu	Gln	Asn	Ala	Thr	Asn	Tyr	Phe	
				100					105					110			
5	Leu	Met	Ser	Leu	Ala	Ile	Ala	Asp	Met	Leu	Leu	Gly	Phe	Leu	Val	Met	
			115					120					125				
	Pro	Val	Ser	Met	Leu	Thr	Ile	Leu	Tyr	Gly	Tyr	Arg	Trp	Pro	Leu	Pro	
		130					135					140					
10	Ser	Lys	Leu	Cys	Ala	Val	Trp	Ile	Tyr	Leu	Asp	Val	Leu	Phe	Ser	Thr	
	145					150					155					160	
	Ala	Ser	Ile	Met	His	Leu	Cys	Ala	Ile	Ser	Leu	Asp	Arg	Tyr	Val	Ala	
				165						170					175		
	Ile	Gln	Asn	Pro	Ile	His	His	Ser	Arg	Phe	Asn	Ser	Arg	Thr	Lys	Ala	
				180					185					190			
15	Phe	Leu	Lys	Ile	Ile	Ala	Val	Trp	Thr	Ile	Ser	Val	Gly	Ile	Ser	Met	
			195					200					205				
	Pro	Ile	Pro	Val	Phe	Gly	Leu	Gln	Asp	Asp	Ser	Lys	Val	Phe	Lys	Glu	
		210					215					220					
20	Gly	Ser	Cys	Leu	Leu	Ala	Asp	Asp	Asn	Phe	Val	Leu	Ile	Gly	Ser	Phe	
	225				230					235						240	
	Val	Ser	Phe	Phe	Ile	Pro	Leu	Thr	Ile	Met	Val	Ile	Thr	Tyr	Phe	Leu	
				245					250					255			
	Thr	Ile	Lys	Ser	Leu	Gln	Lys	Glu	Ala	Thr	Leu	Cys	Val	Ser	Asp	Leu	
			260					265						270			
25	Gly	Thr	Arg	Ala	Lys	Leu	Ala	Ser	Phe	Ser	Phe	Leu	Pro	Gln	Ser	Ser	
		275					280						285				
	Leu	Ser	Ser	Glu	Lys	Leu	Phe	Gln	Arg	Ser	Ile	His	Arg	Glu	Pro	Gly	
		290				295					300						
30	Ser	Tyr	Thr	Gly	Arg	Arg	Thr	Met	Gln	Ser	Ile	Ser	Asn	Glu	Gln	Lys	
	305				310						315					320	
	Ala	Cys	Lys	Val	Leu	Gly	Ile	Val	Phe	Phe	Leu	Phe	Val	Val	Met	Trp	
				325					330					335			
	Cys	Pro	Phe	Phe	Ile	Thr	Asn	Ile	Met	Ala	Val	Ile	Cys	Lys	Glu	Ser	
			340					345					350				
35	Cys	Asn	Glu	Asp	Val	Ile	Gly	Ala	Leu	Leu	Asn	Val	Phe	Val	Trp	Ile	
		355					360					365					
	Gly	Tyr	Leu	Ser	Ser	Ala	Val	Asn	Pro	Leu	Val	Tyr	Thr	Leu	Phe	Asn	

97

370 375 380

Lys Thr Tyr Arg Ser Ala Phe Ser Arg Tyr Ile Gln Cys Gln Tyr Lys
385 390 395 400

5 Glu Asn Lys Lys Pro Leu Gln Leu Ile Leu Val Asn Thr Ile Pro Ala
405 410 415

Leu Ala Tyr Lys Ser Ser Gln Leu Gln Met Gly Gln Lys Lys Asn Ser
420 425 430

Lys Gln Asp Ala Lys Thr Thr Asp Asn Asp Cys Ser Met Val Ala Leu
435 440 445

10 Gly Lys Gln Tyr Ser Glu Glu Ala Ser Lys Asp Asn Ser Asp Gly Val
450 455 460

Asn Glu Lys Val Ser Cys Val
465 470

(124) INFORMATION FOR SEQ ID NO:123:

- 15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 27 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- 20 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:123:

GACCTCGAGG TTGCTTAAGA CTGAAGC

27

(125) INFORMATION FOR SEQ ID NO:124:

- 25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 27 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:124:

ATTCTAGAC ATATGTAGCT TGTACCG

27

(126) INFORMATION FOR SEQ ID NO:125:

- 35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1377 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:125:

	ATGGTGAACC TGAGGAATGC GGTGCATTCA TTCCTTGTGC ACCTAATTGG CCTATTGGTT	60
	TGGCAATGTG ATATTTCTGT GAGCCAGTA GCAGCTATAG TAACTGACAT TTTCAATACC	120
5	TCCGATGGTG GACGCTTCAA ATTCCCAGAC GGGGTACAAA ACTGGCCAGC ACTTTCAATC	180
	GTCATCATAA TAATCATGAC AATAGGTGGC AACATCCTTG TGATCATGGC AGTAAGCATG	240
	GAAAAGAAAC TGCACAATGC CACCAATTAC TTCTTAATGT CCCTAGCCAT TGCTGATATG	300
	CTAGTGGGAC TACTTGTCAT GCCCCTGTCT CTCCTGGCAA TCCTTTATGA TTATGTCTGG	360
	CCACTACCTA GATATTTGTG CCCCGTCTGG ATTTCTTTAG ATGTTTATT TTCAACAGCG	420
10	TCCATCATGC ACCTCTGCGC TATATCGCTG GATCGGTATG TAGCAATACG TAATCCTATT	480
	GAGCATAGCC GTTCAATTC GCGGACTAAG GCCATCATGA AGATTGCTAT TGTTTGGGCA	540
	ATTTCTATAG GTGTATCAGT TCCTATCCCT GTGATTGGAC TGAGGGACGA AGAAAAGGTG	600
	TTCGTGAACA ACACGACGTG CGTGCTCAAC GACCCAAATT TCGTTCCTAT TGGGTCCCTC	660
	GTAGCTTTCT TCATACCGCT GACGATTATG GTGATTACGT ATTGCCTGAC CATCTACGTT	720
15	CTGCGCCGAC AAGCTTTGAT GTTACTGCAC GGCCACACCG AGGAACCGCC TGGACTAAGT	780
	CTGGATTTCC TGAAGTGCTG CAAGAGGAAT ACGGCCGAGG AAGAGAACTC TGCAAACCCCT	840
	AACCAAGACC AGAACGCACG CCGAAGAAAG AAGAAGGAGA GACGTCCTAG GGGCACCATG	900
	CAGGCTATCA ACAATGAAAG AAAAGCTTCG AAAGTCCTTG GGATTGTTTT CTTTGTGTTT	960
	CTGATCATGT GGTGCCCATT TTTCATTACC AATATTCTGT CTGTTCTTTG TGAGAAGTCC	1020
20	TGTAACCAAA AGCTCATGGA AAAGCTTCTG AATGTGTTTG TTTGGATTGG CTATGTTTGT	1080
	TCAGGAATCA ATCCTCTGGT GTATACTCTG TTCAACAAA TTTACCGAAG GGCATTCTCC	1140
	AACTATTTGC GTTGCAATTA TAAGGTAGAG AAAAAGCCTC CTGTCAGGCA GATTCCAAGA	1200
	GTTGCCGCCA CTGCTTTGTC TGGGAGGGAG CTTAATGTTA ACATTTATCG GCATACCAAT	1260
	GAACCGGTGA TCGAGAAAGC CAGTGACAAT GAGCCCGGTA TAGAGATGCA AGTTGAGAAT	1320
25	TTAGAGTTAC CAGTAAATCC CTCCAGTGTG GTTAGCGAAA GGATTAGCAG TGTGTGA	1377

(127) INFORMATION FOR SEQ ID NO:126:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 458 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:126:

5 Met Val Asn Leu Arg Asn Ala Val His Ser Phe Leu Val His Leu Ile
 1 5 10 15
 Gly Leu Leu Val Trp Gln Cys Asp Ile Ser Val Ser Pro Val Ala Ala
 20 25 30
 10 Ile Val Thr Asp Ile Phe Asn Thr Ser Asp Gly Gly Arg Phe Lys Phe
 35 40 45
 Pro Asp Gly Val Gln Asn Trp Pro Ala Leu Ser Ile Val Ile Ile Ile
 50 55 60
 Ile Met Thr Ile Gly Gly Asn Ile Leu Val Ile Met Ala Val Ser Met
 65 70 75 80
 15 Glu Lys Lys Leu His Asn Ala Thr Asn Tyr Phe Leu Met Ser Leu Ala
 85 90 95
 Ile Ala Asp Met Leu Val Gly Leu Leu Val Met Pro Leu Ser Leu Leu
 100 105 110
 20 Ala Ile Leu Tyr Asp Tyr Val Trp Pro Leu Pro Arg Tyr Leu Cys Pro
 115 120 125
 Val Trp Ile Ser Leu Asp Val Leu Phe Ser Thr Ala Ser Ile Met His
 130 135 140
 Leu Cys Ala Ile Ser Leu Asp Arg Tyr Val Ala Ile Arg Asn Pro Ile
 145 150 155 160
 25 Glu His Ser Arg Phe Asn Ser Arg Thr Lys Ala Ile Met Lys Ile Ala
 165 170 175
 Ile Val Trp Ala Ile Ser Ile Gly Val Ser Val Pro Ile Pro Val Ile
 180 185 190
 30 Gly Leu Arg Asp Glu Glu Lys Val Phe Val Asn Asn Thr Thr Cys Val
 195 200 205
 Leu Asn Asp Pro Asn Phe Val Leu Ile Gly Ser Phe Val Ala Phe Phe
 210 215 220
 Ile Pro Leu Thr Ile Met Val Ile Thr Tyr Cys Leu Thr Ile Tyr Val
 225 230 235 240
 35 Leu Arg Arg Gln Ala Leu Met Leu Leu His Gly His Thr Glu Glu Pro
 245 250 255

Pro Gly Leu Ser Leu Asp Phe Leu Lys Cys Cys Lys Arg Asn Thr Ala
260 265 270

Glu Glu Glu Asn Ser Ala Asn Pro Asn Gln Asp Gln Asn Ala Arg Arg
275 280 285

5 Arg Lys Lys Lys Glu Arg Arg Pro Arg Gly Thr Met Gln Ala Ile Asn
290 295 300

Asn Glu Arg Lys Ala Ser Lys Val Leu Gly Ile Val Phe Phe Val Phe
305 310 315 320

10 Leu Ile Met Trp Cys Pro Phe Phe Ile Thr Asn Ile Leu Ser Val Leu
325 330 335

Cys Glu Lys Ser Cys Asn Gln Lys Leu Met Glu Lys Leu Leu Asn Val
340 345 350

Phe Val Trp Ile Gly Tyr Val Cys Ser Gly Ile Asn Pro Leu Val Tyr
355 360 365

15 Thr Leu Phe Asn Lys Ile Tyr Arg Arg Ala Phe Ser Asn Tyr Leu Arg
370 375 380

Cys Asn Tyr Lys Val Glu Lys Lys Pro Pro Val Arg Gln Ile Pro Arg
385 390 395 400

20 Val Ala Ala Thr Ala Leu Ser Gly Arg Glu Leu Asn Val Asn Ile Tyr
405 410 415

Arg His Thr Asn Glu Pro Val Ile Glu Lys Ala Ser Asp Asn Glu Pro
420 425 430

Gly Ile Glu Met Gln Val Glu Asn Leu Glu Leu Pro Val Asn Pro Ser
435 440 445

25 Ser Val Val Ser Glu Arg Ile Ser Ser Val
450 455

(128) INFORMATION FOR SEQ ID NO:127:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:127:

35 GGTAAGCTTG GCAGTCCACG CCAGGCCTTC

30

(129) INFORMATION FOR SEQ ID NO:128:

101

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:128:

TCCGAATTCT CTGTAGACAC AAGGCTTTGG

30

(130) INFORMATION FOR SEQ ID NO:129:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1068 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:129:

ATGGATCAGT TCCCTGAATC AGTGACAGAA AACTTTGAGT ACGATGATTT GGCTGAGGCC 60

TGTTATATTG GGGACATCGT GGTCTTTGGG ACTGTGTTCC TGTCCATATT CTA CTCCGTC 120

ATCTTTGCCA TTGGCCTGGT GGGAAATTTG TTGGTAGTGT TTGCCCTCAC CAACAGCAAG 180

20 AAGCCCAAGA GTGTCACCGA CATTACCTC CTGAACCTGG CCTTGTCTGA TCTGCTGTTT 240

G TAGCCACTT TGCCCTTCTG GACTCACTAT TTGATAAATG AAAAGGGCCT CCACAATGCC 300

ATGTGCAAAT TCACTACCGC CTTCTTCTTC ATCGGCTTTT TTGGAAGCAT ATTCTTCATC 360

ACCGTCATCA GCATTGATAG GTACCTGGCC ATCGTCCTGG CCGCCAACTC CATGAACAAC 420

CGGACCGTGC AGCATGGCGT CACCATCAGC CTAGGCGTCT GGGCAGCAGC CATTTTGGTG 480

25 GCAGCACCCC AGTTCATGTT CACAAAGCAG AAAGAAAATG AATGCCTTGG TGA CTACCCC 540

GAGGTCTCTC AGGAAATCTG GCCCGTGCTC CGCAATGTGG AAACAAATTT TCTTGGCTTC 600

CTACTCCCCC TGCTCATTAT GAGTTATTGC TACTTCAGAA TCATCCAGAC GCTGTTTTC 660

TGCAAGAACC ACAAGAAAGC CAAAGCCATT AA ACTGATCC TTCTGGTGGT CATCGTGTTT 720

TTCCTCTTCT GGACACCCTA CAACGTTATG ATTTTCCTGG AGACGCTTAA GCTCTATGAC 780

30 TTCTTTCCCA GTTGTGACAT GAGGAAGGAT CTGAGGCTGG CCCTCAGTGT GACTGAGACG 840

GTTGCATTTA GCCATTGTTG CCTGAATCCT CTCATCTATG CATTTGCTGG GGAGAAGTTC 900

AGAAGATACC TTTACCACCT GTATGGGAAA TGCCTGGCTG TCCTGTGTGG GCGCTCAGTC 960

CACGTTGATT TCTCCTCATC TGAATCACAA AGGAGCAGGC ATGGAAGTGT TCTGAGCAGC 1020

AATTTTACTT ACCACACGAG TGATGGAGAT GCATTGCTCC TTCTCTGA 1068

(131) INFORMATION FOR SEQ ID NO:130:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 355 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:130:

5	Met Asp Gln Phe Pro Glu Ser Val Thr Glu Asn Phe Glu Tyr Asp Asp	1 5 10 15
	Leu Ala Glu Ala Cys Tyr Ile Gly Asp Ile Val Val Phe Gly Thr Val	20 25 30
15	Phe Leu Ser Ile Phe Tyr Ser Val Ile Phe Ala Ile Gly Leu Val Gly	35 40 45
	Asn Leu Leu Val Val Phe Ala Leu Thr Asn Ser Lys Lys Pro Lys Ser	50 55 60
20	Val Thr Asp Ile Tyr Leu Leu Asn Leu Ala Leu Ser Asp Leu Leu Phe	65 70 75 80
	Val Ala Thr Leu Pro Phe Trp Thr His Tyr Leu Ile Asn Glu Lys Gly	85 90 95
	Leu His Asn Ala Met Cys Lys Phe Thr Thr Ala Phe Phe Phe Ile Gly	100 105 110
25	Phe Phe Gly Ser Ile Phe Phe Ile Thr Val Ile Ser Ile Asp Arg Tyr	115 120 125
	Leu Ala Ile Val Leu Ala Ala Asn Ser Met Asn Asn Arg Thr Val Gln	130 135 140
30	His Gly Val Thr Ile Ser Leu Gly Val Trp Ala Ala Ala Ile Leu Val	145 150 155 160
	Ala Ala Pro Gln Phe Met Phe Thr Lys Gln Lys Glu Asn Glu Cys Leu	165 170 175
	Gly Asp Tyr Pro Glu Val Leu Gln Glu Ile Trp Pro Val Leu Arg Asn	180 185 190
35	Val Glu Thr Asn Phe Leu Gly Phe Leu Leu Pro Leu Leu Ile Met Ser	195 200 205

103

Tyr Cys Tyr Phe Arg Ile Ile Gln Thr Leu Phe Ser Cys Lys Asn His
 210 215 220
 Lys Lys Ala Lys Ala Ile Lys Leu Ile Leu Leu Val Val Ile Val Phe
 225 230 235 240
 5 Phe Leu Phe Trp Thr Pro Tyr Asn Val Met Ile Phe Leu Glu Thr Leu
 245 250 255
 Lys Leu Tyr Asp Phe Phe Pro Ser Cys Asp Met Arg Lys Asp Leu Arg
 260 265 270
 10 Leu Ala Leu Ser Val Thr Glu Thr Val Ala Phe Ser His Cys Cys Leu
 275 280 285
 Asn Pro Leu Ile Tyr Ala Phe Ala Gly Glu Lys Phe Arg Arg Tyr Leu
 290 295 300
 Tyr His Leu Tyr Gly Lys Cys Leu Ala Val Leu Cys Gly Arg Ser Val
 305 310 315 320
 15 His Val Asp Phe Ser Ser Ser Glu Ser Gln Arg Ser Arg His Gly Ser
 325 330 335
 Val Leu Ser Ser Asn Phe Thr Tyr His Thr Ser Asp Gly Asp Ala Leu
 340 345 350
 20 Leu Leu Leu
 355

(132) INFORMATION FOR SEQ ID NO:131:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 32 base pairs
 (B) TYPE: nucleic acid
 25 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: DNA (genomic)
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:131:

GATCTCCAGT AGGCATAAGT GGACAATTCT GG

32

30 (133) INFORMATION FOR SEQ ID NO:132:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 35 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: DNA (genomic)
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:132:

CTCCTTCGGT CCTCCTATCG TTGTCAGAAG

30

(134) INFORMATION FOR SEQ ID NO:133:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:133:

10 AGAAGGCCAA GATCGCGCGG CTGGCCCTCA

30

(135) INFORMATION FOR SEQ ID NO:134:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:134:

CGGCGCCACC GCACGAAAAA GCTCATCTTC

30

20 (136) INFORMATION FOR SEQ ID NO:135:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 33 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:135:

GCCAAGAAGC GGGTGAAGTT CCTGGTGGTG GCA

33

(137) INFORMATION FOR SEQ ID NO:136:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:136:

CAGGCGGAAG GTGAAAGTCC TGGTCCTCGT

30

(138) INFORMATION FOR SEQ ID NO:137:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:137:

CGGCGCCTGC GGGCCAAGCG GCTGGTGGTG GTG

33

(139) INFORMATION FOR SEQ ID NO:138:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:138:

CCAAGCACAA AGCCAAGAAA GTGACCATCA C

31

(140) INFORMATION FOR SEQ ID NO:139:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:139:

GCGCCGGCGC ACCAAATGCT TGCTGGTGGT

30

(141) INFORMATION FOR SEQ ID NO:140:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 41 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:140:

CAAAAAGCTG AAGAAATCTA AGAAGATCAT CTTTATTGTC G

41

(142) INFORMATION FOR SEQ ID NO:141:

5

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:141:

CAAGACCAAG GCAAAACGCA TGATCGCCAT

30

(143) INFORMATION FOR SEQ ID NO:142:

15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:142:

GTCAAGGAGA AGTCCAAAAG GATCATCATC

30

(144) INFORMATION FOR SEQ ID NO:143:

25

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:143:

30

CGCCGCGTGC GGGCCAAGCA GCTCCTGCTC

30

(145) INFORMATION FOR SEQ ID NO:144:

35

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 33 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

107

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:144:

CCTGATAAGC GCTATAAAAT GGTCCTGTTT CGA

33

(146) INFORMATION FOR SEQ ID NO:145:

5

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 36 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:145:

GAAAGACAAA AGAGAGTCAA GAGGATGTCT TTATTG

36

(147) INFORMATION FOR SEQ ID NO:146:

15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 33 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:146:

CGGAGAAAGA GGGTGAAACG CACAGCCATC GCC

33

(148) INFORMATION FOR SEQ ID NO:147:

25

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:147:

30

AAGCTTCAGC GGGCCAAGGC ACTGGTCACC

30

(149) INFORMATION FOR SEQ ID NO:148:

35

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:148:

CAGCGGCAGA AGGC¹AAAAG GGTGGCCATC

30

(150) INFORMATION FOR SEQ ID NO:149:

5

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:149:

CGGCAGAAGG CGAAGCGCAT GATCCTCGCG

30

(151) INFORMATION FOR SEQ ID NO:150:

15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:150:

GAGCGCAACA AGGCCAAAAA GGTGATCATC

30

(152) INFORMATION FOR SEQ ID NO:151:

25

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 39 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:151:

30

GGTGTAACA AAAAGGCTAA AAACACAATT ATTCTTATT

39

(153) INFORMATION FOR SEQ ID NO:152:

35

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:152:

GAGAGCCAGC TCAAGAGCAC CGTGGTG

27

(154) INFORMATION FOR SEQ ID NO:153:

- 5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:153:

CCACAAGCAA ACCAAGAAAA TGCTGGCTGT

30

(155) INFORMATION FOR SEQ ID NO:154:

- 15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:154:

CATCAAGTGT ATCATGTGCC AAGTACGCCC

30

(156) INFORMATION FOR SEQ ID NO:155:

- 25 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:155:

30 CTAGAGAGTC AGATGAAGTG TACAGTAGTG GCAC

34

(157) INFORMATION FOR SEQ ID NO:156:

- 35 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:156:

CGGACAAAAG TGAAAACTAA AAAGATGTTC CTCATT

36

(158) INFORMATION FOR SEQ ID NO:157:

- 5 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 33 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:157:

GCTGAGGTTC GCAATAAACT AACCATGTTT GTG

33

(159) INFORMATION FOR SEQ ID NO:158:

- 15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 29 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:158:

GGGAGGCCGA GCTGAAAGCC ACCCTGCTC

29

(160) INFORMATION FOR SEQ ID NO:159:

- 25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 31 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:159:

30 CAAGATCAAG AGAGCCAAAA CCTTCATCAT G

31

(161) INFORMATION FOR SEQ ID NO:160:

- 35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 31 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

111

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:160:

CCGGAGACAA GTGAAGAGA TGCTGTTTGT C

31

(162) INFORMATION FOR SEQ ID NO:161:

- 5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:161:

GCAAGGACCA GATCAAGCGG CTGGTGCTCA

30

(163) INFORMATION FOR SEQ ID NO:162:

- 15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:162:

CAAGAAAGCC AAAGCCAAGA AACTGATCCT TCTG

34

(164) INFORMATION FOR SEQ ID NO:163:

- 25 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1068 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:163:

30 ATGGAAGATT TGGAGGAAAC ATTATTTGAA GAATTTGAAA ACTATTCCTA TGACCTAGAC 60

TATTACTCTC TGGAGTCTGA TTTGGAGGAG AAAGTCCAGC TGGGAGTTGT TCACTGGGTC 120

TCCCTGGTGT TATATTGTTT GGCTTTTGT CTGGGAATTC CAGGAAATGC CATCGTCATT 180

TGGTTCACGG GGCTCAAGTG GAAGAAGACA GTCACCACTC TGTGGTTCCT CAATCTAGCC 240

ATTGCGGATT TCATTTTCT TCTCTTCTG CCCCTGTACA TCTCCTATGT GGCCATGAAT 300

TTCCACTGGC CCTTTGGCAT CTGGCTGTGC AAAGCCAATT CCTTCACTGC CCAGTTGAAC 360
 ATGTTTGCCA GTGTTTTTTT CCTGACAGTG ATCAGCCTGG ACCACTATAT CCACTTGATC 420
 CATCCTGTCT TATCTCATCG GCATCGAACC CTCAAGAACT CTCTGATTGT CATTATATTC 480
 ATCTGGCTTT TGGCTTCTCT AATTGGCGGT CCTGCCCTGT ACTTCCGGGA CACTGTGGAG 540
 5 TTCAATAATC ATACTCTTTG CTATAACAAT TTTCAGAAGC ATGATCCTGA CCTCACTTTG 600
 ATCAGGCACC ATGTTCTGAC TTGGGTGAAA TTTATCATTG GCTATCTCTT CCCTTTGCTA 660
 ACAATGAGTA TTTGCTACTT GTGTCTCATC TTCAAGGTGA AGAAGCGAAC AGTCCTGATC 720
 TCCAGTAGGC ATAAGTGGAC AATTCTGGTT GTGGTTGTGG CCTTTGTGGT TTGCTGGACT 780
 CCTTATCACC TGTTTAGCAT TTGGGAGCTC ACCATTCACC ACAATAGCTA TTCCACCAT 840
 10 GTGATGCAGG CTGGAATCCC CCTCTCCACT GGTTCGGCAT TCCTCAATAG TTGCTTGAAC 900
 CCCATCCTTT ATGTCCTAAT TAGTAAGAAG TTCCAAGCTC GCTTCCGGTC CTCAGTTGCT 960
 GAGATACTCA AGTACACACT GTGGGAAGTC AGCTGTTCTG GCACAGTGAG TGAACAGCTC 1020
 AGGAACTCAG AAACCAAGAA TCTGTGTCTC CTGGAAACAG CTCAATAA 1068

(165) INFORMATION FOR SEQ ID NO:164:

15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 355 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant
 20 (ii) MOLECULE TYPE: protein
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:164:
 Met Glu Asp Leu Glu Glu Thr Leu Phe Glu Glu Phe Glu Asn Tyr Ser
 1 5 10 15
 Tyr Asp Leu Asp Tyr Tyr Ser Leu Glu Ser Asp Leu Glu Glu Lys Val
 25 20 25 30
 Gln Leu Gly Val Val His Trp Val Ser Leu Val Leu Tyr Cys Leu Ala
 35 40 45
 Phe Val Leu Gly Ile Pro Gly Asn Ala Ile Val Ile Trp Phe Thr Gly
 50 55 60
 30 Leu Lys Trp Lys Lys Thr Val Thr Thr Leu Trp Phe Leu Asn Leu Ala
 65 70 75 80
 Ile Ala Asp Phe Ile Phe Leu Leu Phe Leu Pro Leu Tyr Ile Ser Tyr
 85 90 95

113

	Val	Ala	Met	Asn	Phe	His	Trp	Pro	Phe	Gly	Ile	Trp	Leu	Cys	Lys	Ala	
				100					105					110			
	Asn	Ser	Phe	Thr	Ala	Gln	Leu	Asn	Met	Phe	Ala	Ser	Val	Phe	Phe	Leu	
			115					120					125				
5	Thr	Val	Ile	Ser	Leu	Asp	His	Tyr	Ile	His	Leu	Ile	His	Pro	Val	Leu	
		130					135					140					
	Ser	His	Arg	His	Arg	Thr	Leu	Lys	Asn	Ser	Leu	Ile	Val	Ile	Ile	Phe	
	145					150					155					160	
10	Ile	Trp	Leu	Leu	Ala	Ser	Leu	Ile	Gly	Gly	Pro	Ala	Leu	Tyr	Phe	Arg	
					165					170					175		
	Asp	Thr	Val	Glu	Phe	Asn	Asn	His	Thr	Leu	Cys	Tyr	Asn	Asn	Phe	Gln	
				180					185					190			
	Lys	His	Asp	Pro	Asp	Leu	Thr	Leu	Ile	Arg	His	His	Val	Leu	Thr	Trp	
			195					200					205				
15	Val	Lys	Phe	Ile	Ile	Gly	Tyr	Leu	Phe	Pro	Leu	Leu	Thr	Met	Ser	Ile	
		210				215						220					
	Cys	Tyr	Leu	Cys	Leu	Ile	Phe	Lys	Val	Lys	Lys	Arg	Thr	Val	Leu	Ile	
	225				230						235					240	
20	Ser	Ser	Arg	His	Lys	Trp	Thr	Ile	Leu	Val	Val	Val	Val	Ala	Phe	Val	
				245						250					255		
	Val	Cys	Trp	Thr	Pro	Tyr	His	Leu	Phe	Ser	Ile	Trp	Glu	Leu	Thr	Ile	
				260					265					270			
	His	His	Asn	Ser	Tyr	Ser	His	His	Val	Met	Gln	Ala	Gly	Ile	Pro	Leu	
			275					280					285				
25	Ser	Thr	Gly	Leu	Ala	Phe	Leu	Asn	Ser	Cys	Leu	Asn	Pro	Ile	Leu	Tyr	
		290					295					300					
	Val	Leu	Ile	Ser	Lys	Lys	Phe	Gln	Ala	Arg	Phe	Arg	Ser	Ser	Val	Ala	
	305				310					315						320	
30	Glu	Ile	Leu	Lys	Tyr	Thr	Leu	Trp	Glu	Val	Ser	Cys	Ser	Gly	Thr	Val	
				325						330				335			
	Ser	Glu	Gln	Leu	Arg	Asn	Ser	Glu	Thr	Lys	Asn	Leu	Cys	Leu	Leu	Glu	
			340						345				350				
	Thr	Ala	Gln														
			355														

35 (166) INFORMATION FOR SEQ ID NO:165:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1089 base pairs

(B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:165:

	ATGGGCAACC ACACGTGGGA GGGCTGCCAC GTGGACTCGC GCGTGGACCA CCTCTTTCCG	60
	CCATCCCTCT ACATCTTTGT CATCGGCGTG GGGCTGCCCA CCAACTGCCT GGCTCTGTGG	120
	GCGGCCTACC GCCAGGTGCA ACAGCGCAAC GAGCTGGGCG TCTACCTGAT GAACCTCAGC	180
	ATCGCCGACC TGCTGTACAT CTGCACGCTG CCGCTGTGGG TGGACTACTT CCTGCACCAC	240
10	GACAACTGGA TCCACGCCCC CGGGTCCTGC AAGCTCTTTG GGTTCATCTT CTACACCAAT	300
	ATCTACATCA GCATCGCCTT CCTGTGTGTC ATCTCGGTGG ACCGCTACCT GGCTGTGGCC	360
	CACCCACTCC GCTTCGCCCC CCTGCGCCGC GTCAAGACCG CCGTGGCCGT GAGCTCCGTG	420
	GTCTGGGCCA CGGAGCTGGG CGCCAACTCG GCGCCCCTGT TCCATGACGA GCTCTTCCGA	480
	GACCGCTACA ACCACACCTT CTGCTTTGAG AAGTTCCCCA TGAAGGCTG GGTGGCCTGG	540
15	ATGAACCTCT ATCGGGTGTT CGTGGGCTTC CTCTTCCCGT GGGCGCTCAT GCTGCTGTGC	600
	TACCGGGGCA TCCTGCGGGC CGTGCGGGGC AGCGTGTCCA CCGAGCGCCA GGAGAAGGCC	660
	AAGATCGCGC GGCTGGCCCT CAGCCTCATC GCCATCGTGC TGGTCTGCTT TGCGCCCTAT	720
	CACGTGCTCT TGCTGTCCCG CAGCGCCATC TACCTGGGCC GCCCCTGGGA CTGCGGCTTC	780
	GAGGAGCGCG TCTTTTCTGC ATACCACAGC TCACTGGCTT TCACCAGCCT CAACTGTGTG	840
20	GCGGACCCCA TCCTCTACTG CCTGGTCAAC GAGGGCGCCC GCAGCGATGT GGCCAAGGCC	900
	CTGCACAACC TGCTCCGCTT TCTGGCCAGC GACAAGCCCC AGGAGATGGC CAATGCCTCG	960
	CTCACCCTGG AGACCCCACT CACCTCCAAG AGGAACAGCA CAGCCAAAGC CATGACTGGC	1020
	AGCTGGGCGG CCACTCCGCC TTCCAGGGG GACCAGGTGC AGCTGAAGAT GCTGCCGCCA	1080
	GCACAATGA	1089

25 (167) INFORMATION FOR SEQ ID NO:166:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 362 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

30

(ii) MOLECULE TYPE: protein

115

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:166:

	Met	Gly	Asn	His	Thr	Trp	Glu	Gly	Cys	His	Val	Asp	Ser	Arg	Val	Asp	
	1				5					10					15		
5	His	Leu	Phe	Pro	Pro	Ser	Leu	Tyr	Ile	Phe	Val	Ile	Gly	Val	Gly	Leu	
				20					25				30				
	Pro	Thr	Asn	Cys	Leu	Ala	Leu	Trp	Ala	Ala	Tyr	Arg	Gln	Val	Gln	Gln	
			35					40					45				
	Arg	Asn	Glu	Leu	Gly	Val	Tyr	Leu	Met	Asn	Leu	Ser	Ile	Ala	Asp	Leu	
			50				55					60					
10	Leu	Tyr	Ile	Cys	Thr	Leu	Pro	Leu	Trp	Val	Asp	Tyr	Phe	Leu	His	His	
	65					70					75				80		
	Asp	Asn	Trp	Ile	His	Gly	Pro	Gly	Ser	Cys	Lys	Leu	Phe	Gly	Phe	Ile	
					85					90					95		
15	Phe	Tyr	Thr	Asn	Ile	Tyr	Ile	Ser	Ile	Ala	Phe	Leu	Cys	Cys	Ile	Ser	
				100					105					110			
	Val	Asp	Arg	Tyr	Leu	Ala	Val	Ala	His	Pro	Leu	Arg	Phe	Ala	Arg	Leu	
			115					120					125				
	Arg	Arg	Val	Lys	Thr	Ala	Val	Ala	Val	Ser	Ser	Val	Val	Trp	Ala	Thr	
			130				135						140				
20	Glu	Leu	Gly	Ala	Asn	Ser	Ala	Pro	Leu	Phe	His	Asp	Glu	Leu	Phe	Arg	
	145					150					155					160	
	Asp	Arg	Tyr	Asn	His	Thr	Phe	Cys	Phe	Glu	Lys	Phe	Pro	Met	Glu	Gly	
					165					170					175		
25	Trp	Val	Ala	Trp	Met	Asn	Leu	Tyr	Arg	Val	Phe	Val	Gly	Phe	Leu	Phe	
				180					185					190			
	Pro	Trp	Ala	Leu	Met	Leu	Leu	Ser	Tyr	Arg	Gly	Ile	Leu	Arg	Ala	Val	
			195					200					205				
	Arg	Gly	Ser	Val	Ser	Thr	Glu	Arg	Gln	Glu	Lys	Ala	Lys	Ile	Ala	Arg	
		210					215					220					
30	Leu	Ala	Leu	Ser	Leu	Ile	Ala	Ile	Val	Leu	Val	Cys	Phe	Ala	Pro	Tyr	
	225					230					235					240	
	His	Val	Leu	Leu	Leu	Ser	Arg	Ser	Ala	Ile	Tyr	Leu	Gly	Arg	Pro	Trp	
					245					250					255		
35	Asp	Cys	Gly	Phe	Glu	Glu	Arg	Val	Phe	Ser	Ala	Tyr	His	Ser	Ser	Leu	
				260					265					270			
	Ala	Phe	Thr	Ser	Leu	Asn	Cys	Val	Ala	Asp	Pro	Ile	Leu	Tyr	Cys	Leu	

275

280

285

Val Asn Glu Gly Ala Arg Ser Asp Val Ala Lys Ala Leu His Asn Leu
 290 295 300

5 Leu Arg Phe Leu Ala Ser Asp Lys Pro Gln Glu Met Ala Asn Ala Ser
 305 310 315 320

Leu Thr Leu Glu Thr Pro Leu Thr Ser Lys Arg Asn Ser Thr Ala Lys
 325 330 335

Ala Met Thr Gly Ser Trp Ala Ala Thr Pro Pro Ser Gln Gly Asp Gln
 340 345 350

10 Val Gln Leu Lys Met Leu Pro Pro Ala Gln
 355 360

(168) INFORMATION FOR SEQ ID NO:167:

(i) SEQUENCE CHARACTERISTICS:

15 (A) LENGTH: 1002 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:167:

20 ATGGAGTCCT CAGGCAACCC AGAGAGCACC ACCTTTTTTTT ACTATGACCT TCAGAGCCAG 60
 CCGTGTGAGA ACCAGGCCTG GGTCTTTGCT ACCCTCGCCA CCACTGTCCT GTACTGCCTG 120
 GTGTTTCTCC TCAGCCTAGT GGGCAACAGC CTGGTCCTGT GGGTCCTGGT GAAGTATGAG 180
 AGCCTGGAGT CCCTACCAA CATCTTCATC CTCAACCTGT GCCTCTCAGA CCTGGTGTTC 240
 GCCTGCTTGT TGCCTGTGTG GATCTCCCCA TACCACTGGG GCTGGGTGCT GGGAGACTTC 300
 25 CTCTGCAAAC TCCTCAATAT GATCTTCTCC ATCAGCCTCT ACAGCAGCAT CTTCTTCCTG 360
 ACCATCATGA CCATCCACCG CTACCTGTG GTAGTGAGCC CCCTCTCCAC CCTGCGCGTC 420
 CCCACCCTCC GCTGCCGGGT GCTGGTGACC ATGGCTGTGT GGGTAGCCAG CATCCTGTCC 480
 TCCATCCTCG ACACCATCTT CCACAAGGTG CTTTCTTCGG GCTGTGATTA TTCCGAATC 540
 ACGTGGTACC TCACCTCCGT CTACCAGCAC AACCTCTTCT TCCTGCTGTC CCTGGGGATT 600
 30 ATCCTGTTCT GCTACGTGGA GATCCTCAGG ACCCTGTTCC GCTCACGCTC CAAGCGGCGC 660
 CACCGCACGA AAAAGCTCAT CTTCGCCATC GTGGTGGCCT ACTTCCTCAG CTGGGGTCCC 720
 TACAACCTCA CCCTGTTTCT GCAGACGCTG TTTCGGACCC AGATCATCCG GAGCTGCGAG 780

117

GCCAAACAGC AGCTAGAATA CGCCCTGCTC ATCTGCCGCA ACCTCGCCTT CTCCCCTGCTC 840
 TGCTTTAACC CGGTGCTCTA TGTCTTCGTG GGGGTCAAGT TCCGCACACA CCTGAAACAT 900
 GTTCTCCGGC AGTTCTGGTT CTGCCGGCTG CAGGCACCCA GCCCAGCCTC GATCCCCCAG 960
 TCCCCTGGTG CCTTCGCCTA TGAGGGCGCC TCCTTCTACT GA 1002

5 (169) INFORMATION FOR SEQ ID NO:168:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 333 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

10 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:168:

	Met	Glu	Ser	Ser	Gly	Asn	Pro	Glu	Ser	Thr	Thr	Phe	Phe	Tyr	Tyr	Asp	
	1				5					10					15		
15	Leu	Gln	Ser	Gln	Pro	Cys	Glu	Asn	Gln	Ala	Trp	Val	Phe	Ala	Thr	Leu	
				20					25					30			
	Ala	Thr	Thr	Val	Leu	Tyr	Cys	Leu	Val	Phe	Leu	Leu	Ser	Leu	Val	Gly	
				35				40					45				
20	Asn	Ser	Leu	Val	Leu	Trp	Val	Leu	Val	Lys	Tyr	Glu	Ser	Leu	Glu	Ser	
		50					55					60					
	Leu	Thr	Asn	Ile	Phe	Ile	Leu	Asn	Leu	Cys	Leu	Ser	Asp	Leu	Val	Phe	
	65					70					75					80	
	Ala	Cys	Leu	Leu	Pro	Val	Trp	Ile	Ser	Pro	Tyr	His	Trp	Gly	Trp	Val	
					85					90					95		
25	Leu	Gly	Asp	Phe	Leu	Cys	Lys	Leu	Leu	Asn	Met	Ile	Phe	Ser	Ile	Ser	
			100						105					110			
	Leu	Tyr	Ser	Ser	Ile	Phe	Phe	Leu	Thr	Ile	Met	Thr	Ile	His	Arg	Tyr	
			115					120					125				
30	Leu	Ser	Val	Val	Ser	Pro	Leu	Ser	Thr	Leu	Arg	Val	Pro	Thr	Leu	Arg	
		130					135					140					
	Cys	Arg	Val	Leu	Val	Thr	Met	Ala	Val	Trp	Val	Ala	Ser	Ile	Leu	Ser	
	145					150				155					160		
	Ser	Ile	Leu	Asp	Thr	Ile	Phe	His	Lys	Val	Leu	Ser	Ser	Gly	Cys	Asp	
				165					170					175			
35	Tyr	Ser	Glu	Leu	Thr	Trp	Tyr	Leu	Thr	Ser	Val	Tyr	Gln	His	Asn	Leu	
				180					185					190			

Phe Phe Leu Leu Ser Leu Gly Ile Ile Leu Phe Cys Tyr Val Glu Ile
 195 200 205
 Leu Arg Thr Leu Phe Arg Ser Arg Ser Lys Arg Arg His Arg Thr Lys
 210 215 220
 5 Lys Leu Ile Phe Ala Ile Val Val Ala Tyr Phe Leu Ser Trp Gly Pro
 225 230 235 240
 Tyr Asn Phe Thr Leu Phe Leu Gln Thr Leu Phe Arg Thr Gln Ile Ile
 245 250 255
 10 Arg Ser Cys Glu Ala Lys Gln Gln Leu Glu Tyr Ala Leu Leu Ile Cys
 260 265 270
 Arg Asn Leu Ala Phe Ser His Cys Cys Phe Asn Pro Val Leu Tyr Val
 275 280 285
 Phe Val Gly Val Lys Phe Arg Thr His Leu Lys His Val Leu Arg Gln
 290 295 300
 15 Phe Trp Phe Cys Arg Leu Gln Ala Pro Ser Pro Ala Ser Ile Pro His
 305 310 315 320
 Ser Pro Gly Ala Phe Ala Tyr Glu Gly Ala Ser Phe Tyr
 325 330

(170) INFORMATION FOR SEQ ID NO:169:

- 20 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 987 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- 25 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:169:

ATGGACAACG CCTCGTTCTC GGAGCCCTGG CCCGCCAACG CATCGGGCCC GGACCCGGCG 60
 CTGAGCTGCT CCAACGCGTC GACTCTGGCG CCGCTGCCGG CGCCGCTGGC GGTGGCTGTA 120
 CCAGTTGTCT ACGCGGTGAT CTGCGCCGTG GGTCTGGCGG GCAACTCCGC CGTGCTGTAC 180
 30 GTGTTGCTGC GGGCGCCCCG CATGAAGACC GTCACCAACC TGTTCATCCT CAACCTGGCC 240
 ATCGCCGACG AGCTCTTCAC GCTGGTGCTG CCCATCAACA TCGCCGACTT CCTGCTGCGG 300
 CAGTGGCCCT TCGGGGAGCT CATGTGCAAG CTCATCGTGG CTATCGACCA GTACAACACC 360
 TTCTCCAGCC TCTACTTCCT CACCGTCATG AGCGCCGACC GCTACCTGGT GGTGTTGGCC 420
 ACTGCGGAGT CGCGCCGGGT GGCCGGCCGC ACCTACAGCG CCGCGCGCGC GGTGAGCCTG 480

119

GCCGTGTGGG GGATCGTCAC ACTCGTCGTG CTGCCCTTCG CAGTCTTCGC CCGGCTAGAC 540
 GACGAGCAGG GCCGGCGCCA GTGCGTGCTA GTCTTTCCGC AGCCCGAGGC CTTCTGGTGG 600
 CGCGCGAGCC GCCTCTACAC GCTCGTGCTG GGCTTCGCCA TCCCCGTGTC CACCATCTGT 660
 GTCCTCTATA CCACCCTGCT GTGCCGGCTG CATGCCATGC GGCTGGACAG CCACGCCAAG 720
 5 GCCCTGGAGC GCGCCAAGAA GCGGGTGAAG TTCCTGGTGG TGGCAATCCT GGCGGTGTGC 780
 CTCCTCTGCT GGACGCCCTA CCACCTGAGC ACCGTGGTGG CGCTCACCAC CGACCTCCCG 840
 CAGACGCCGC TGGTCATCGC TATCTCCTAC TTCATCACCA GCCTGACGTA CGCCAACAGC 900
 TGCCTCAACC CCTTCCTCTA CGCCTTCCTG GACGCCAGCT TCCGCAGGAA CCTCCGCCAG 960
 CTGATAACTT GCCGCGCGGC AGCCTGA 987

10 (171) INFORMATION FOR SEQ ID NO:170:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 328 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - 15 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:170:

Met Asp Asn Ala Ser Phe Ser Glu Pro Trp Pro Ala Asn Ala Ser Gly
 1 5 10 15
 20 Pro Asp Pro Ala Leu Ser Cys Ser Asn Ala Ser Thr Leu Ala Pro Leu
 20 25 30
 Pro Ala Pro Leu Ala Val Ala Val Pro Val Val Tyr Ala Val Ile Cys
 35 40 45
 25 Ala Val Gly Leu Ala Gly Asn Ser Ala Val Leu Tyr Val Leu Leu Arg
 50 55 60
 Ala Pro Arg Met Lys Thr Val Thr Asn Leu Phe Ile Leu Asn Leu Ala
 65 70 75 80
 Ile Ala Asp Glu Leu Phe Thr Leu Val Leu Pro Ile Asn Ile Ala Asp
 85 90 95
 30 Phe Leu Leu Arg Gln Trp Pro Phe Gly Glu Leu Met Cys Lys Leu Ile
 100 105 110
 Val Ala Ile Asp Gln Tyr Asn Thr Phe Ser Ser Leu Tyr Phe Leu Thr
 115 120 125
 Val Met Ser Ala Asp Arg Tyr Leu Val Val Leu Ala Thr Ala Glu Ser

130 135 140

Arg Arg Val Ala Gly Arg Thr Tyr Ser Ala Ala Arg Ala Val Ser Leu
145 150 155 160

5 Ala Val Trp Gly Ile Val Thr Leu Val Val Leu Pro Phe Ala Val Phe
 165 170 175

Ala Arg Leu Asp Asp Glu Gln Gly Arg Arg Gln Cys Val Leu Val Phe
 180 185 190

Pro Gln Pro Glu Ala Phe Trp Trp Arg Ala Ser Arg Leu Tyr Thr Leu
 195 200 205

10 Val Leu Gly Phe Ala Ile Pro Val Ser Thr Ile Cys Val Leu Tyr Thr
 210 215 220

Thr Leu Leu Cys Arg Leu His Ala Met Arg Leu Asp Ser His Ala Lys
225 230 235 240

15 Ala Leu Glu Arg Ala Lys Lys Arg Val Lys Phe Leu Val Val Ala Ile
 245 250 255

Leu Ala Val Cys Leu Leu Cys Trp Thr Pro Tyr His Leu Ser Thr Val
 260 265 270

Val Ala Leu Thr Thr Asp Leu Pro Gln Thr Pro Leu Val Ile Ala Ile
 275 280 285

20 Ser Tyr Phe Ile Thr Ser Leu Thr Tyr Ala Asn Ser Cys Leu Asn Pro
 290 295 300

Phe Leu Tyr Ala Phe Leu Asp Ala Ser Phe Arg Arg Asn Leu Arg Gln
305 310 315 320

25 Leu Ile Thr Cys Arg Ala Ala Ala
 325

(172) INFORMATION FOR SEQ ID NO:171:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1002 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:171:

ATGCAGGCCG CTGGGCACCC AGAGCCCCTT GACAGCAGGG GCTCCTTCTC CCTCCCCACG 60

35 ATGGGTGCCA ACGTCTCTCA GGACAATGGC ACTGGCCACA ATGCCACCTT CTCCGAGCCA 120

CTGCCGTTCC TCTATGTGCT CCTGCCCGCC GTGTACTCCG GGATCTGTGC TGTGGGGCTG 180

121

ACTGGCAACA CGGCCGTCAT CCTTGTAATC CTAAGGGCGC CCAAGATGAA GACGGTGACC 240
 AACGTGTTCA TCCTGAACCT GGCCGTCGCC GACGGGCTCT TCACGCTGGT ACTGCCTGTC 300
 AACATCGCGG AGCACCTGCT GCAGTACTGG CCCTTCGGGG AGCTGCTCTG CAAGCTGGTG 360
 CTGGCCGTCG ACCACTACAA CATCTTCTCC AGCATCTACT TCCTAGCCGT GATGAGCGTG 420
 5 GACCGATACC TGGTGGTGCT GGCCACCGTG AGGTCCCGCC ACATGCCCTG GCGCACCTAC 480
 CGGGGGGCGA AGGTCGCCAG CCTGTGTGTC TGGCTGGGCG TCACGGTCCT GGTTC TGCCC 540
 TTCTTCTCTT TCGCTGGCGT CTACAGCAAC GAGCTGCAGG TCCAAGCTG TGGGCTGAGC 600
 TTCCCGTGGC CCGAGCAGGT CTGGTTCAAG GCCAGCCGTG TCTACACGTT GGTCTGGGC 660
 TTCGTGCTGC CCGTGTGCAC CATCTGTGTG CTCTACACAG ACCTCCTGCG CAGGCTGCGG 720
 10 GCCGTGCGGC TCCGCTCTGG AGCCAAGGCT CTAGGCAAGG CCAGGCGGAA GGTGAAAGTC 780
 CTGGTCCTCG TCGTGCTGGC CGTGTGCCTC CTCTGCTGGA CGCCCTTCCA CCTGGCCTCT 840
 GTCGTGGCCC TGACCACGGA CCTGCCCCAG ACCCCACTGG TCATCAGTAT GTCCTACGTC 900
 ATCACCAGCC TCACGTACGC CAACTCGTGC CTGAACCCCT TCCTCTACGC CTTTCTAGAT 960
 GACAACCTCC GGAAGAACTT CCGCAGCATA TTGCGGTGCT GA 1002

15 (173) INFORMATION FOR SEQ ID NO:172:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 333 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

20 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:172:

Met Gln Ala Ala Gly His Pro Glu Pro Leu Asp Ser Arg Gly Ser Phe
1 5 10 15

25 Ser Leu Pro Thr Met Gly Ala Asn Val Ser Gln Asp Asn Gly Thr Gly
20 25 30

His Asn Ala Thr Phe Ser Glu Pro Leu Pro Phe Leu Tyr Val Leu Leu
35 40 45

30 Pro Ala Val Tyr Ser Gly Ile Cys Ala Val Gly Leu Thr Gly Asn Thr
50 55 60

Ala Val Ile Leu Val Ile Leu Arg Ala Pro Lys Met Lys Thr Val Thr
65 70 75 80

	Asn	Val	Phe	Ile	Leu	Asn	Leu	Ala	Val	Ala	Asp	Gly	Leu	Phe	Thr	Leu	
					85					90					95		
	Val	Leu	Pro	Val	Asn	Ile	Ala	Glu	His	Leu	Leu	Gln	Tyr	Trp	Pro	Phe	
					100					105					110		
5	Gly	Glu	Leu	Leu	Cys	Lys	Leu	Val	Leu	Ala	Val	Asp	His	Tyr	Asn	Ile	
					115					120					125		
	Phe	Ser	Ser	Ile	Tyr	Phe	Leu	Ala	Val	Met	Ser	Val	Asp	Arg	Tyr	Leu	
					130					135					140		
10	Val	Val	Leu	Ala	Thr	Val	Arg	Ser	Arg	His	Met	Pro	Trp	Arg	Thr	Tyr	
					145					150					155		
	Arg	Gly	Ala	Lys	Val	Ala	Ser	Leu	Cys	Val	Trp	Leu	Gly	Val	Thr	Val	
					165					170					175		
	Leu	Val	Leu	Pro	Phe	Phe	Ser	Phe	Ala	Gly	Val	Tyr	Ser	Asn	Glu	Leu	
					180					185					190		
15	Gln	Val	Pro	Ser	Cys	Gly	Leu	Ser	Phe	Pro	Trp	Pro	Glu	Gln	Val	Trp	
					195					200					205		
	Phe	Lys	Ala	Ser	Arg	Val	Tyr	Thr	Leu	Val	Leu	Gly	Phe	Val	Leu	Pro	
					210					215					220		
20	Val	Cys	Thr	Ile	Cys	Val	Leu	Tyr	Thr	Asp	Leu	Leu	Arg	Arg	Leu	Arg	
					225					230					235		
	Ala	Val	Arg	Leu	Arg	Ser	Gly	Ala	Lys	Ala	Leu	Gly	Lys	Ala	Arg	Arg	
					245					250					255		
	Lys	Val	Lys	Val	Leu	Val	Leu	Val	Val	Leu	Ala	Val	Cys	Leu	Leu	Cys	
					260					265					270		
25	Trp	Thr	Pro	Phe	His	Leu	Ala	Ser	Val	Val	Ala	Leu	Thr	Thr	Asp	Leu	
					275					280					285		
	Pro	Gln	Thr	Pro	Leu	Val	Ile	Ser	Met	Ser	Tyr	Val	Ile	Thr	Ser	Leu	
					290					295					300		
30	Thr	Tyr	Ala	Asn	Ser	Cys	Leu	Asn	Pro	Phe	Leu	Tyr	Ala	Phe	Leu	Asp	
					305					310					315		
	Asp	Asn	Phe	Arg	Lys	Asn	Phe	Arg	Ser	Ile	Leu	Arg	Cys				
					325					330							

(174) INFORMATION FOR SEQ ID NO:173:

- 35 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1107 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:173:

```

ATGGTCCTTG AGGTGAGTGA CCACCAAGTG CTAAATGACG CCGAGGTTGC CGCCCTCCTG   60
GAGAACTTCA GCTCTTCCTA TGACTATGGA GAAAACGAGA GTGACTCGTG CTGTACCTCC   120
5 CCGCCCTGCC CACAGGACTT CAGCCTGAAC TTCGACCGGG CCTTCCTGCC AGCCCTCTAC   180
AGCCTCCTCT TTCTGCTGGG GCTGCTGGGC AACGGCGCGG TGGCAGCCGT GCTGCTGAGC   240
CGGCGGACAG CCCTGAGCAG CACCGACACC TTCCTGCTCC ACCTAGCTGT AGCAGACACG   300
CTGCTGGTGC TGACACTGCC GCTCTGGGCA GTGGACGCTG CCGTCCAGTG GGTCTTTGGC   360
TCTGGCCTCT GCAAAGTGGC AGGTGCCCTC TTCAACATCA ACTTCTACGC AGGAGCCCTC   420
10 CTGCTGGCCT GCATCAGCTT TGACCGCTAC CTGAACATAG TTCATGCCAC CCAGCTCTAC   480
CGCCGGGGGC CCCC GGCCCG CGTGACCCTC ACCTGCCTGG CTGTCTGGGG GCTCTGCCTG   540
CTTTTCGCCC TCCCAGACTT CATCTTCCTG TCGGCCCACC ACGACGAGCG CCTCAACGCC   600
ACCCACTGCC AATACAACCT CCCACAGGTG GGCCGCACGG CTCTGCGGGT GCTGCAGCTG   660
GTGGCTGGCT TTCTGCTGCC CCTGCTGGTC ATGGCCTACT GCTATGCCCA CATCCTGGCC   720
15 GTGCTGCTGG TTTCCAGGGG CCAGCGGCGC CTGCGGGCCA AGCGGCTGGT GGTGGTGGTC   780
GTGGTGGCCT TTGCCCTCTG CTGGACCCCC TATCACCTGG TGGTGCTGGT GGACATCCTC   840
ATGGACCTGG GCGCTTTGGC CCGCAACTGT GGCCGAGAAA GCAGGGTAGA CGTGGCCAAG   900
TCGGTCACCT CAGGCCTGGG CTACATGCAC TGCTGCCTCA ACCCGCTGCT CTATGCCTTT   960
GTAGGGGTCA AGTTCCGGGA GCGGATGTGG ATGCTGCTCT TGCGCCTGGG CTGCCCCAAC  1020
20 CAGAGAGGGC TCCAGAGGCA GCCATCGTCT TCCCGCCGGG ATTCATCCTG GTCTGAGACC  1080
TCAGAGGCCT CCTACTCGGG CTTGTGA                                     1107

```

(175) INFORMATION FOR SEQ ID NO:174:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 368 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:174:

Met Val Leu Glu Val Ser Asp His Gln Val Leu Asn Asp Ala Glu Val
1 5 10 15

Ala Ala Leu Leu Glu Asn Phe Ser Ser Ser Tyr Asp Tyr Gly Glu Asn
20 25 30

5 Glu Ser Asp Ser Cys Cys Thr Ser Pro Pro Cys Pro Gln Asp Phe Ser
35 40 45

Leu Asn Phe Asp Arg Ala Phe Leu Pro Ala Leu Tyr Ser Leu Leu Phe
50 55 60

10 Leu Leu Gly Leu Leu Gly Asn Gly Ala Val Ala Ala Val Leu Leu Ser
65 70 75 80

Arg Arg Thr Ala Leu Ser Ser Thr Asp Thr Phe Leu Leu His Leu Ala
85 90 95

Val Ala Asp Thr Leu Leu Val Leu Thr Leu Pro Leu Trp Ala Val Asp
100 105 110

15 Ala Ala Val Gln Trp Val Phe Gly Ser Gly Leu Cys Lys Val Ala Gly
115 120 125

Ala Leu Phe Asn Ile Asn Phe Tyr Ala Gly Ala Leu Leu Leu Ala Cys
130 135 140

20 Ile Ser Phe Asp Arg Tyr Leu Asn Ile Val His Ala Thr Gln Leu Tyr
145 150 155 160

Arg Arg Gly Pro Pro Ala Arg Val Thr Leu Thr Cys Leu Ala Val Trp
165 170 175

Gly Leu Cys Leu Leu Phe Ala Leu Pro Asp Phe Ile Phe Leu Ser Ala
180 185 190

25 His His Asp Glu Arg Leu Asn Ala Thr His Cys Gln Tyr Asn Phe Pro
195 200 205

Gln Val Gly Arg Thr Ala Leu Arg Val Leu Gln Leu Val Ala Gly Phe
210 215 220

30 Leu Leu Pro Leu Leu Val Met Ala Tyr Cys Tyr Ala His Ile Leu Ala
225 230 235 240

Val Leu Leu Val Ser Arg Gly Gln Arg Arg Leu Arg Ala Lys Arg Leu
245 250 255

Val Val Val Val Val Val Ala Phe Ala Leu Cys Trp Thr Pro Tyr His
260 265 270

35 Leu Val Val Leu Val Asp Ile Leu Met Asp Leu Gly Ala Leu Ala Arg
275 280 285

Asn Cys Gly Arg Glu Ser Arg Val Asp Val Ala Lys Ser Val Thr Ser

125

290 295 300

Gly Leu Gly Tyr Met His Cys Cys Leu Asn Pro Leu Leu Tyr Ala Phe
305 310 315 320

5 Val Gly Val Lys Phe Arg Glu Arg Met Trp Met Leu Leu Leu Arg Leu
325 330 335

Gly Cys Pro Asn Gln Arg Gly Leu Gln Arg Gln Pro Ser Ser Ser Arg
340 345 350

Arg Asp Ser Ser Trp Ser Glu Thr Ser Glu Ala Ser Tyr Ser Gly Leu
355 360 365

10 (176) INFORMATION FOR SEQ ID NO:175:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1074 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

15 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:175:

ATGGCTGATG ACTATGGCTC TGAATCCACA TCTTCCATGG AAGACTACGT TAACTTCAAC 60

TTCAGTGAAT TCTACTGTGA GAAAAACAAT GTCAGGCAGT TTGCGAGCCA TTTCTCCCA 120

20 CCCTTGTAAT GGCTCGTGTT CATCGTGGGT GCCTTGGGCA ACAGTCTTGT TATCCTTGTC 180

TACTGGTACT GCACAAGAGT GAAGACCATG ACCGACATGT TCCTTTTGAA TTTGGCAATT 240

GCTGACCTCC TCTTTCTTGT CACTCTTCCC TTCTGGGCCA TTGCTGCTGC TGACCAAGTGG 300

AAGTTCCAGA CCTTCATGTG CAAGGTGGTC AACAGCATGT ACAAGATGAA CTTCTACAGC 360

TGTGTGTTGC TGATCATGTG CATCAGCGTG GACAGGTACA TTGCCATTGC CCAGGCCATG 420

25 AGAGCACATA CTTGGAGGGA GAAAAGGCTT TTGTACAGCA AAATGGTTTG CTTTACCATC 480

TGGGTATTGG CAGCTGCTCT CTGCATCCCA GAAATCTTAT ACAGCCAAAT CAAGGAGGAA 540

TCCGGCATTG CTATCTGCAC CATGGTTTAC CCTAGCGATG AGAGCACCAA ACTGAAGTCA 600

GCTGTCTTGA CCCTGAAGGT CATTCTGGGG TTCTTCCTTC CTTTCGTGGT CATGGCTTGC 660

TGCTATACCA TCATCATTCA CACCCTGATA CAAGCCAAGA AGTCTTCCAA GCACAAAGCC 720

30 AAGAAAGTGA CCATCACTGT CCTGACCGTC TTTGTCTTGT CTCAGTTTCC CTACAACTGC 780

ATTTTGTGGG TGCAGACCAT TGACGCCTAT GCCATGTTCA TCTCCAAGTG TGCCGTTTCC 840

ACCAACATTG ACATCTGCTT CCAGGTCACC CAGACCATCG CTTTCTTCCA CAGTTGCCTG 900
 AACCTGTTC TCTATGTTTT TGTGGGTGAG AGATTCCGCC GGGATCTCGT GAAAACCCTG 960
 AAGAACTTGG GTTGCATCAG CCAGGCCAG TGGGTTTCAT TTACAAGGAG AGAGGGAAGC 1020
 TTGAAGCTGT CGTCTATGTT GCTGGAGACA ACCTCAGGAG CACTCTCCCT CTGA 1074

5 (177) INFORMATION FOR SEQ ID NO:176:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 357 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

10

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:176:

Met Ala Asp Asp Tyr Gly Ser Glu Ser Thr Ser Ser Met Glu Asp Tyr
 1 5 10 15
 Val Asn Phe Asn Phe Thr Asp Phe Tyr Cys Glu Lys Asn Asn Val Arg
 20 25 30
 Gln Phe Ala Ser His Phe Leu Pro Pro Leu Tyr Trp Leu Val Phe Ile
 35 40 45
 Val Gly Ala Leu Gly Asn Ser Leu Val Ile Leu Val Tyr Trp Tyr Cys
 50 55 60
 Thr Arg Val Lys Thr Met Thr Asp Met Phe Leu Leu Asn Leu Ala Ile
 65 70 75 80
 Ala Asp Leu Leu Phe Leu Val Thr Leu Pro Phe Trp Ala Ile Ala Ala
 85 90 95
 Ala Asp Gln Trp Lys Phe Gln Thr Phe Met Cys Lys Val Val Asn Ser
 100 105 110
 Met Tyr Lys Met Asn Phe Tyr Ser Cys Val Leu Leu Ile Met Cys Ile
 115 120 125
 Ser Val Asp Arg Tyr Ile Ala Ile Ala Gln Ala Met Arg Ala His Thr
 130 135 140
 Trp Arg Glu Lys Arg Leu Leu Tyr Ser Lys Met Val Cys Phe Thr Ile
 145 150 155 160
 Trp Val Leu Ala Ala Ala Leu Cys Ile Pro Glu Ile Leu Tyr Ser Gln
 165 170 175
 Ile Lys Glu Glu Ser Gly Ile Ala Ile Cys Thr Met Val Tyr Pro Ser
 180 185 190

127

Asp Glu Ser Thr Lys Leu Lys Ser Ala Val Leu Thr Leu Lys Val Ile
 195 200 205
 Leu Gly Phe Phe Leu Pro Phe Val Val Met Ala Cys Cys Tyr Thr Ile
 210 215 220
 5 Ile Ile His Thr Leu Ile Gln Ala Lys Lys Ser Ser Lys His Lys Ala
 225 230 235 240
 Lys Lys Val Thr Ile Thr Val Leu Thr Val Phe Val Leu Ser Gln Phe
 245 250 255
 10 Pro Tyr Asn Cys Ile Leu Leu Val Gln Thr Ile Asp Ala Tyr Ala Met
 260 265 270
 Phe Ile Ser Asn Cys Ala Val Ser Thr Asn Ile Asp Ile Cys Phe Gln
 275 280 285
 Val Thr Gln Thr Ile Ala Phe Phe His Ser Cys Leu Asn Pro Val Leu
 290 295 300
 15 Tyr Val Phe Val Gly Glu Arg Phe Arg Arg Asp Leu Val Lys Thr Leu
 305 310 315 320
 Lys Asn Leu Gly Cys Ile Ser Gln Ala Gln Trp Val Ser Phe Thr Arg
 325 330 335
 20 Arg Glu Gly Ser Leu Lys Leu Ser Ser Met Leu Leu Glu Thr Thr Ser
 340 345 350
 Gly Ala Leu Ser Leu
 355

(178) INFORMATION FOR SEQ ID NO:177:

25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1110 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:177:

ATGGCCTCAT CGACCACTCG GGGCCCCAGG GTTCTGACT TATTTTCTGG GCTGCCGCCG 60
 GCGGTCACAA CTCCCGCCAA CCAGAGCGCA GAGGCCTCGG CGGGCAACGG GTCGGTGGCT 120
 GGCGCGGACG CTCCAGCCGT CACGCCCTTC CAGAGCCTGC AGCTGGTGCA TCAGCTGAAG 180
 GGGCTGATCG TGCTGCTCTA CAGCGTCGTG GTGGTCGTGG GGCTGGTGGG CAACTGCCTG 240
 35 CTGGTGCTGG TGATCGCGCG GGTGCCGCGG CTGCACAACG TGACGAACTT CCTCATCGGC 300

```

AACCTGGCCT TGTCCGACGT GTCATGTGC ACCGCCTGCG TGCCGCTCAC GCTGGCCTAT 360
GCCTTCGAGC CACGCGGCTG GGTGTTTCGGC GGC GGCCCTGT GCCACCTGGT CTTCTTCCTG 420
CAGCCGGTCA CCGTCTATGT GTCGGTGTTC ACGCTCACCA CCATCSCAGT GGACCGCTAC 480
GTCGTGCTGG TGCACCCGCT GAGGCGCGCA TCTCGCTGCG CCTCAGCCTA CGCTGTGCTG 540
5 GCCATCTGGG CGCTGTCCGC GGTGCTGGCG CTGCCGCCCG CCGTGCCACAC CTATCACGTG 600
GAGCTCAAGC CGCAGGACGT GCGCCTCTGC GAGGAGTTCT GGGGCTCCCA GGAGCGCCAG 660
CGCCAGCTCT ACGCCTGGGG GCTGCTGCTG GTCACCTACC TGCTCCCTCT GCTGGTCATC 720
CTCCTGTCTT ACGTCCGGGT GTCAGTGAAG CTCCGCAACC GCGTGGTGCC GGGCTGCGTG 780
ACCCAGAGCC AGGCCGACTG GGACCGCGCT CGGCGCCGGC GCACCAAATG CTTGCTGGTG 840
10 GTGGTCGTGG TGGTGTTCGC CGTCTGCTGG CTGCCGCTGC ACGTCTTCAA CCTGCTGCGG 900
GACCTCGACC CCCACGCCAT CGACCCTTAC GCCTTTGGGC TGGTGCAGCT GCTCTGCCAC 960
TGGCTCGCCA TGAGTTCGGC CTGCTACAAC CCCTTCATCT ACGCCTGGCT GCACGACAGC 1020
TTCCGCGAGG AGCTGCGCAA ACTGTTGGTC GCTTGGCCCC GCAAGATAGC CCCCCATGGC 1080
CAGAATATGA CCGTCAGCGT GGTCATCTGA 1110

```

15 (179) INFORMATION FOR SEQ ID NO:178:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 369 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

20

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:178:

```

Met Ala Ser Ser Thr Thr Arg Gly Pro Arg Val Ser Asp Leu Phe Ser
1          5          10          15
25 Gly Leu Pro Pro Ala Val Thr Thr Pro Ala Asn Gln Ser Ala Glu Ala
    20          25          30
Ser Ala Gly Asn Gly Ser Val Ala Gly Ala Asp Ala Pro Ala Val Thr
    35          40          45
30 Pro Phe Gln Ser Leu Gln Leu Val His Gln Leu Lys Gly Leu Ile Val
    50          55          60
Leu Leu Tyr Ser Val Val Val Val Gly Leu Val Gly Asn Cys Leu
65          70          75          80

```

[illegible]

(180) INFORMATION FOR SEQ ID NO:179:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1083 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:179:

```

10  ATGGACCCAG AAGAACTTC AGTTATTTG GATTATTACT ATGCTACGAG CCCAACTCT   60
    GACATCAGGG AGACCCACTC CCATGTTCTT TACACCTCTG TCTTCCTTCC AGTCTTTTAC   120
    ACAGCTGTGT TCCTGACTGG AGTGCTGGGG AACCTTGTTT TCATGGGAGC GTTGCAATTC   180
    AAACCCGGCA GCCGAAGACT GATCGACATC TTTATCATCA ATCTGGCTGC CTCTGACTTC   240
    ATTTTCTCTG TCACATTGCC TCTCTGGGTG GATAAAGAAG CATCTCTAGG ACTGTGGAGG   300
15  ACGGGCTCCT TCCTGTGCAA AGGGAGCTCC TACATGATCT CCGTCAATAT GCACTGCAGT   360
    GTCCTCCTGC TCACTTGCAT GAGTGTTGAC CGCTACCTGG CCATTGTGTG GCCAGTCGTA   420
    TCCAGGAAAT TCAGAAGGAC AGACTGTGCA TATGTAGTCT GTGCCAGCAT CTGGTTTATC   480
    TCCTGCCTGC TGGGGTTGCC TACTCTTCTG TCCAGGGAGC TCACGCTGAT TGATGATAAG   540
    CCATACTGTG CAGAGAAAAA GGCAACTCCA ATTAACTCA TATGGTCCCT GGTGGCCTTA   600
20  ATTTTCACCT TTTTGTCCC TTGTTGAGC ATTGTGACCT GCTACTGTTG CATTGCAAGG   660
    AAGCTGTGTG CCCATTACCA GCAATCAGGA AAGCACAACA AAAAGCTGAA GAAATCTAAG   720
    AAGATCATCT TTATTGTCGT GGCAGCCTTT CTTGTCTCCT GGCTGCCCTT CAATACTTTC   780
    AAGTTCCTGG CCATTGTCTC TGGGTTGCGG CAAGAACACT ATTTACCCTC AGCTATTCTT   840
    CAGCTTGATA TGGAGGTGAG TGGACCCTTG GCATTTGCCA ACAGCTGTGT CAACCCTTTC   900
25  ATTTACTATA TCTTCGACAG CTACATCCGC CGGGCCATTG TCCACTGCTT GTGCCCTTGC   960
    CTGAAAAACT ATGACTTTGG GAGTAGCACT GAGACATCAG ATAGTCACCT CACTAAGGCT  1020
    CTCTCCACCT TCATTCATGC AGAAGATTTT GCCAGGAGGA GGAAGAGGTC TGTGTCACTC  1080
    TAA
                                         1083

```

(181) INFORMATION FOR SEQ ID NO:180:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 360 amino acids

131

(B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:180:

	Met	Asp	Pro	Glu	Glu	Thr	Ser	Val	Tyr	Leu	Asp	Tyr	Tyr	Tyr	Ala	Thr	
	1				5					10					15		
	Ser	Pro	Asn	Ser	Asp	Ile	Arg	Glu	Thr	His	Ser	His	Val	Pro	Tyr	Thr	
			20						25					30			
10	Ser	Val	Phe	Leu	Pro	Val	Phe	Tyr	Thr	Ala	Val	Phe	Leu	Thr	Gly	Val	
			35					40					45				
	Leu	Gly	Asn	Leu	Val	Leu	Met	Gly	Ala	Leu	His	Phe	Lys	Pro	Gly	Ser	
		50					55					60					
	Arg	Arg	Leu	Ile	Asp	Ile	Phe	Ile	Ile	Asn	Leu	Ala	Ala	Ser	Asp	Phe	
15	65					70				75						80	
	Ile	Phe	Leu	Val	Thr	Leu	Pro	Leu	Trp	Val	Asp	Lys	Glu	Ala	Ser	Leu	
					85					90					95		
	Gly	Leu	Trp	Arg	Thr	Gly	Ser	Phe	Leu	Cys	Lys	Gly	Ser	Ser	Tyr	Met	
				100					105					110			
20	Ile	Ser	Val	Asn	Met	His	Cys	Ser	Val	Leu	Leu	Leu	Thr	Cys	Met	Ser	
			115					120					125				
	Val	Asp	Arg	Tyr	Leu	Ala	Ile	Val	Trp	Pro	Val	Val	Ser	Arg	Lys	Phe	
		130					135					140					
	Arg	Arg	Thr	Asp	Cys	Ala	Tyr	Val	Val	Cys	Ala	Ser	Ile	Trp	Phe	Ile	
25	145					150				155						160	
	Ser	Cys	Leu	Leu	Gly	Leu	Pro	Thr	Leu	Leu	Ser	Arg	Glu	Leu	Thr	Leu	
					165				170						175		
	Ile	Asp	Asp	Lys	Pro	Tyr	Cys	Ala	Glu	Lys	Lys	Ala	Thr	Pro	Ile	Lys	
				180					185					190			
30	Leu	Ile	Trp	Ser	Leu	Val	Ala	Leu	Ile	Phe	Thr	Phe	Phe	Val	Pro	Leu	
			195					200					205				
	Leu	Ser	Ile	Val	Thr	Cys	Tyr	Cys	Cys	Ile	Ala	Arg	Lys	Leu	Cys	Ala	
		210					215					220					
	His	Tyr	Gln	Gln	Ser	Gly	Lys	His	Asn	Lys	Lys	Leu	Lys	Lys	Ser	Lys	
35	225					230				235						240	
	Lys	Ile	Ile	Phe	Ile	Val	Val	Ala	Ala	Phe	Leu	Val	Ser	Trp	Leu	Pro	
					245					250					255		

Phe Asn Thr Phe Lys Phe Leu Ala Ile Val Ser Gly Leu Arg Gln Glu
 260 265 270
 His Tyr Leu Pro Ser Ala Ile Leu Gln Leu Gly Met Glu Val Ser Gly
 275 280 285
 5 Pro Leu Ala Phe Ala Asn Ser Cys Val Asn Pro Phe Ile Tyr Tyr Ile
 290 295 300
 Phe Asp Ser Tyr Ile Arg Arg Ala Ile Val His Cys Leu Cys Pro Cys
 305 310 315 320
 10 Leu Lys Asn Tyr Asp Phe Gly Ser Ser Thr Glu Thr Ser Asp Ser His
 325 330 335
 Leu Thr Lys Ala Leu Ser Thr Phe Ile His Ala Glu Asp Phe Ala Arg
 340 345 350
 Arg Arg Lys Arg Ser Val Ser Leu
 355 360

15 (182) INFORMATION FOR SEQ ID NO:181:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1020 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 20 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:181:

ATGAATGGCC TTGAAGTGGC TCCCCAGGT CTGATCACCA ACTTCTCCCT GGCCACGGCA 60
 GAGCAATGTG GCCAGGAGAC GCCACTGGAG AACATGCTGT TCGCCTCCTT CTACCTTCTG 120
 25 GATTTTATCC TGGCTTTAGT TGGCAATACC CTGGCTCTGT GGCTTTTCAT CCGAGACCAC 180
 AAGTCCGGGA CCCC GGCCAA CGTGTTCTTG ATGCATCTGG CCGTGGCCGA CTTGTCGTGC 240
 GTGCTGGTCC TGCCACCCG CCTGGTCTAC CACTTCTCTG GGAACCACTG GCCATTTGGG 300
 GAAATCGCAT GCCGTCTCAC CGGCTTCCTC TTCTACCTCA ACATGTACGC CAGCATCTAC 360
 TTCCTCACCT GCATCAGCGC CGACCGTTTC CTGGCCATTG TGCACCCGGT CAAGTCCCTC 420
 30 AAGCTCCGCA GGCCCCTCTA CGCACACCTG GCCTGTGCCT TCCTGTGGGT GGTGGTGGCT 480
 GTGGCCATGG CCCC GCTGCT GGTGAGCCCA CAGACCGTGC AGACCAACCA CACGGTGGTC 540
 TGCCTGCAGC TGTACCGGGA GAAGGCCTCC CACCATGCCC TGGTGTCCCT GGCAGTGGCC 600
 TTCACCTTCC CGTTCATCAC CACGGTCACC TGCTACCTGC TGATCATCCG CAGCCTGCGG 660

133

CAGGGCCTGC GTGTGGAGAA GCGCCTCAAG ACCAAGGCAA AACGCATGAT CGCCATAGTG 720
 CTGGCCATCT TCCTGGTCTG CTTCGTGCCC TACCACGTCA ACCGCTCCGT CTACGTGCTG 780
 CACTACCGCA GCCATGGGGC CTCCTGCGCC ACCCAGCGCA TCCTGGCCCT GGCAAACCGC 840
 ATCACCTCCT GCCTCACCAG CCTCAACGGG GCACTCGACC CCATCATGTA TTTCTTCGTG 900
 5 GCTGAGAAGT TCCGCCACGC CCTGTGCAAC TTGCTCTGTG GCAAAAGGCT CAAGGGCCCCG 960
 CCCCCAGCT TCGAAGGGAA AACCAACGAG AGCTCGCTGA GTGCCAAGTC AGAGCTGTGA 1020

(183) INFORMATION FOR SEQ ID NO:182:

(i) SEQUENCE CHARACTERISTICS:

10

- (A) LENGTH: 339 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:182:

15

Met Asn Gly Leu Glu Val Ala Pro Pro Gly Leu Ile Thr Asn Phe Ser
 1 5 10 15

Leu Ala Thr Ala Glu Gln Cys Gly Gln Glu Thr Pro Leu Glu Asn Met
 20 25 30

20

Leu Phe Ala Ser Phe Tyr Leu Leu Asp Phe Ile Leu Ala Leu Val Gly
 35 40 45

Asn Thr Leu Ala Leu Trp Leu Phe Ile Arg Asp His Lys Ser Gly Thr
 50 55 60

Pro Ala Asn Val Phe Leu Met His Leu Ala Val Ala Asp Leu Ser Cys
 65 70 75 80

25

Val Leu Val Leu Pro Thr Arg Leu Val Tyr His Phe Ser Gly Asn His
 85 90 95

Trp Pro Phe Gly Glu Ile Ala Cys Arg Leu Thr Gly Phe Leu Phe Tyr
 100 105 110

30

Leu Asn Met Tyr Ala Ser Ile Tyr Phe Leu Thr Cys Ile Ser Ala Asp
 115 120 125

Arg Phe Leu Ala Ile Val His Pro Val Lys Ser Leu Lys Leu Arg Arg
 130 135 140

Pro Leu Tyr Ala His Leu Ala Cys Ala Phe Leu Trp Val Val Val Ala
 145 150 155 160

35

Val Ala Met Ala Pro Leu Leu Val Ser Pro Gln Thr Val Gln Thr Asn

165 170 175

His Thr Val Val Cys Leu Gln Leu Tyr Arg Glu Lys Ala Ser His His
180 185 190

5 Ala Leu Val Ser Leu Ala Val Ala Phe Thr Phe Pro Phe Ile Thr Thr
195 200 205

Val Thr Cys Tyr Leu Leu Ile Ile Arg Ser Leu Arg Gln Gly Leu Arg
210 215 220

Val Glu Lys Arg Leu Lys Thr Lys Ala Lys Arg Met Ile Ala Ile Val
225 230 235 240

10 Leu Ala Ile Phe Leu Val Cys Phe Val Pro Tyr His Val Asn Arg Ser
245 250 255

Val Tyr Val Leu His Tyr Arg Ser His Gly Ala Ser Cys Ala Thr Gln
260 265 270

15 Arg Ile Leu Ala Leu Ala Asn Arg Ile Thr Ser Cys Leu Thr Ser Leu
275 280 285

Asn Gly Ala Leu Asp Pro Ile Met Tyr Phe Phe Val Ala Glu Lys Phe
290 295 300

Arg His Ala Leu Cys Asn Leu Leu Cys Gly Lys Arg Leu Lys Gly Pro
305 310 315 320

20 Pro Pro Ser Phe Glu Gly Lys Thr Asn Glu Ser Ser Leu Ser Ala Lys
325 330 335

Ser Glu Leu

(183) INFORMATION FOR SEQ ID NO:183:

- 25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 996 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- 30 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:183:

ATGATCACCC TGAACAATCA AGATCAACCT GTCCCTTTTA ACAGCTCACA TCCAGATGAA 60

TACAAAATTG CAGCCCTTGT CTTCTATAGC TGTATCTTCA TAATTGGATT ATTTGTTAAC 120

ATCACTGCAT TATGGGTTTT CAGTTGTACC ACCAAGAAGA GAACCACGGT AACCATCTAT 180

35 ATGATGAATG TGGCATTAGT GGACTTGATA TTTATAATGA CTTTACCCTT TCGAATGTTT 240

135

TATTATGCAA AAGATGAATG GCCATTTGGA GAGTACTTCT GCCAGATTCT TGGAGCTCTC 300
 ACAGTGT TTTT ACCCAAGCAT TGCTTTATGG CTTCTTGCCT TTATTAGTGC TGACAGATAC 360
 ATGGCCATTG TACAGCCGAA GTACGCCAAA GAACTTAAAA ACACGTGCAA AGCCGTGCTG 420
 GCGTGTGTGG GAGTCTGGAT AATGACCCTG ACCACGACCA CCCCTCTGCT ACTGCTCTAT 480
 5 AAAGACCCAG ATAAAGACTC CACTCCCGCC ACCTGCCTCA AGATTTCTGA CATCATCTAT 540
 CTAAAAGCTG TGAACGTGCT GAACCTCACT CGACTGACAT TTTTTTCTT GATTCCTTTG 600
 TTCATCATGA TTGGGTGCTA CTTGGTCATT ATTCATAATC TCCTTCACGG CAGGACGTCT 660
 AAGCTGAAAC CCAAAGTCAA GGAGAAGTCC AAAAGGATCA TCATCACGCT GCTGGTGACG 720
 GTGCTCGTCT GCTTTATGCC CTTCCACATC TGTTTCGCTT TCCTGATGCT GGGAACGGGG 780
 10 GAGAATAGTT ACAATCCCTG GGGAGCCTTT ACCACCTTCC TCATGAACCT CAGCACGTGT 840
 CTGGATGTGA TTCTCTACTA CATCGTTTCA AAACAATTC AGGCTCGAGT CATTAGTGTC 900
 ATGCTATACC GTAATTACCT TCGAAGCATG CGCAGAAAAA GTTCCGATC TGGTAGTCTA 960
 AGGTCATAA GCAATATAAA CAGTGAAATG TTATGA 996

(185) INFORMATION FOR SEQ ID NO:184:

15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 331 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

20 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:184:

Met	Ile	Thr	Leu	Asn	Asn	Gln	Asp	Gln	Pro	Val	Pro	Phe	Asn	Ser	Ser
1				5				10					15		
His	Pro	Asp	Glu	Tyr	Lys	Ile	Ala	Ala	Leu	Val	Phe	Tyr	Ser	Cys	Ile
25			20				25					30			
Phe	Ile	Ile	Gly	Leu	Phe	Val	Asn	Ile	Thr	Ala	Leu	Trp	Val	Phe	Ser
	35					40					45				
Cys	Thr	Thr	Lys	Lys	Arg	Thr	Val	Thr	Ile	Tyr	Met	Met	Asn	Val	
	50				55				60						
Ala	Leu	Val	Asp	Leu	Ile	Phe	Ile	Met	Thr	Leu	Pro	Phe	Arg	Met	Phe
65				70				75					80		
Tyr	Tyr	Ala	Lys	Asp	Glu	Trp	Pro	Phe	Gly	Glu	Tyr	Phe	Cys	Gln	Ile
			85					90					95		

	Leu Gly Ala Leu Thr Val Phe Tyr Pro Ser Ile Ala Leu Trp Leu Leu	
	100	110
	Ala Phe Ile Ser Ala Asp Arg Tyr Met Ala Ile Val Gln Pro Lys Tyr	
	115	125
5	Ala Lys Glu Leu Lys Asn Thr Cys Lys Ala Val Leu Ala Cys Val Gly	
	130	140
	Val Trp Ile Met Thr Leu Thr Thr Thr Thr Pro Leu Leu Leu Tyr	
	145	160
10	Lys Asp Pro Asp Lys Asp Ser Thr Pro Ala Thr Cys Leu Lys Ile Ser	
	165	175
	Asp Ile Ile Tyr Leu Lys Ala Val Asn Val Leu Asn Leu Thr Arg Leu	
	180	190
	Thr Phe Phe Phe Leu Ile Pro Leu Phe Ile Met Ile Gly Cys Tyr Leu	
	195	205
15	Val Ile Ile His Asn Leu Leu His Gly Arg Thr Ser Lys Leu Lys Pro	
	210	220
	Lys Val Lys Glu Lys Ser Lys Arg Ile Ile Ile Thr Leu Leu Val Gln	
	225	240
20	Val Leu Val Cys Phe Met Pro Phe His Ile Cys Phe Ala Phe Leu Met	
	245	255
	Leu Gly Thr Gly Glu Asn Ser Tyr Asn Pro Trp Gly Ala Phe Thr Thr	
	260	270
	Phe Leu Met Asn Leu Ser Thr Cys Leu Asp Val Ile Leu Tyr Tyr Ile	
	275	285
25	Val Ser Lys Gln Phe Gln Ala Arg Val Ile Ser Val Met Leu Tyr Arg	
	290	300
	Asn Tyr Leu Arg Ser Met Arg Arg Lys Ser Phe Arg Ser Gly Ser Leu	
	305	320
30	Arg Ser Leu Ser Asn Ile Asn Ser Glu Met Leu	
	325	330

(186) INFORMATION FOR SEQ ID NO:185:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1077 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:185:

ATGCCCTCTG TGTCTCCAGC GGGGCCCTCG GCCGGGGCAG TCCCCAATGC CACCGCAGTG 60
 ACAACAGTGC GGACCAATGC CAGCGGGCTG GAGGTGCCCC TGTTCACCT GTTTGCCCCG 120
 CTGGACGAGG AGCTGCATGG CACCTTCCCA GGCCTGTGCG TGGCGCTGAT GGCGGTGCAC 180
 5 GGAGCCATCT TCCTGGCAGG GCTGGTGCTC AACGGGCTGG CGCTGTACGT CTTCTGCTGC 240
 CGCACCCGGG CCAAGACACC CTCAGTCATC TACACCATCA ACCTGGTGGT GACCGATCTA 300
 CTGGTAGGGC TGTCCTGCC CACGCGCTTC GCTGTGTACT ACGGCGCCAG GGGCTGCCTG 360
 CGCTGTGCCT TCCCGCACGT CCTCGGTAC TTCCTCAACA TGCACTGCTC CATCCTCTTC 420
 CTCACCTGCA TCTGCGTGGA CCGCTACCTG GCCATCGTGC GGCCGAAGG CTCCGCGCGC 480
 10 TGCCGCCAGC CTGCCTGTGC CAGGGCCGTG TGCGCCTTCG TGTGGCTGGC CGCCGGTGCC 540
 GTCACCTGT CGGTGCTGGG CGTGACAGG AGCCGGCCCT GCTGCCGTGT CTTTGCGCTG 600
 ACTGTCCTGG AGTTCCTGCT GCCCCTGCTG GTCATCAGCG TGTTTACCGG CCGCATCATG 660
 TGTGCACTGT CGCGGCCGGG TCTGCTCCAC CAGGGTCGCC AGCGCCGCGT GCGGGCCAAG 720
 CAGCTCCTGC TCACGGTGCT CATCATCTTT CTCGTCTGCT TCACGCCCTT CCACGCCCGC 780
 15 CAAGTGGCCG TGGCGCTGTG GCCCGACATG CCACACCACA CGAGCCTCGT GGTCTACCAC 840
 GTGGCCGTGA CCCTCAGCAG CCTCAACAGC TGCATGGACC CCATCGTCTA CTGCTTCGTC 900
 ACCAGTGGCT TCCAGGCCAC CGTCCGAGGC CTCTTCGGCC AGCACGGAGA GCGTGAGCCC 960
 AGCAGCGGTG ACGTGGTCAG CATGCACAGG AGCTCCAAGG GCTCAGGCCG TCATCACATC 1020
 CTCAGTGCCG GCCCTCACGC CCTCACCCAG GCCCTGGCTA ATGGGCCCCGA GGCTTAG 1077

20 (187) INFORMATION FOR SEQ ID NO:186:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 358 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 25 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:186:

Met Pro Ser Val Ser Pro Ala Gly Pro Ser Ala Gly Ala Val Pro Asn
 1 5 10 15
 30 Ala Thr Ala Val Thr Thr Val Arg Thr Asn Ala Ser Gly Leu Glu Val
 20 25 30

Pro Leu Phe His Leu Phe Ala Arg Leu Asp Glu Glu Leu His Gly Thr
 35 40 45
 Phe Pro Gly Leu Cys Val Ala Leu Met Ala Val His Gly Ala Ile Phe
 50 55 60
 5 Leu Ala Gly Leu Val Leu Asn Gly Leu Ala Leu Tyr Val Phe Cys Cys
 65 70 75 80
 Arg Thr Arg Ala Lys Thr Pro Ser Val Ile Tyr Thr Ile Asn Leu Val
 85 90 95
 10 Val Thr Asp Leu Leu Val Gly Leu Ser Leu Pro Thr Arg Phe Ala Val
 100 105 110
 Tyr Tyr Gly Ala Arg Gly Cys Leu Arg Cys Ala Phe Pro His Val Leu
 115 120 125
 Gly Tyr Phe Leu Asn Met His Cys Ser Ile Leu Phe Leu Thr Cys Ile
 130 135 140
 15 Cys Val Asp Arg Tyr Leu Ala Ile Val Arg Pro Glu Gly Ser Arg Ala
 145 150 155 160
 Cys Arg Gln Pro Ala Cys Ala Arg Ala Val Cys Ala Phe Val Trp Leu
 165 170 175
 20 Ala Ala Gly Ala Val Thr Leu Ser Val Leu Gly Val Thr Gly Ser Arg
 180 185 190
 Pro Cys Cys Arg Val Phe Ala Leu Thr Val Leu Glu Phe Leu Leu Pro
 195 200 205
 Leu Leu Val Ile Ser Val Phe Thr Gly Arg Ile Met Cys Ala Leu Ser
 210 215 220
 25 Arg Pro Gly Leu Leu His Gln Gly Arg Gln Arg Arg Val Arg Ala Lys
 225 230 235 240
 Gln Leu Leu Leu Thr Val Leu Ile Ile Phe Leu Val Cys Phe Thr Pro
 245 250 255
 30 Phe His Ala Arg Gln Val Ala Val Ala Leu Trp Pro Asp Met Pro His
 260 265 270
 His Thr Ser Leu Val Val Tyr His Val Ala Val Thr Leu Ser Ser Leu
 275 280 285
 Asn Ser Cys Met Asp Pro Ile Val Tyr Cys Phe Val Thr Ser Gly Phe
 290 295 300
 35 Gln Ala Thr Val Arg Gly Leu Phe Gly Gln His Gly Glu Arg Glu Pro
 305 310 315 320
 Ser Ser Gly Asp Val Val Ser Met His Arg Ser Ser Lys Gly Ser Gly

325 330 335

Arg His His Ile Leu Ser Ala Gly Pro His Ala Leu Thr Gln Ala Leu
 340 345 350

Ala Asn Gly Pro Glu Ala
 355

(188) INFORMATION FOR SEQ ID NO:187:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1050 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:187:

ATGAACTCCA CCTTGGATGG TAATCAGAGC AGCCACCCTT TTTGCCTCTT GGCATTTGGC 60

15 TATTTGGAAA CTGTCAATTT TTGCCTTTTG GAAGTATTGA TTATTGTCTT TCTAACTGTA 120

TTGATTATTT CTGGCAACAT CATTGTGATT TTTGTATTTC ACTGTGCACC TTTGTTGAAC 180

CATCACACTA CAAGTTATTT TATCCAGACT ATGGCATATG CTGACCTTTT TGTTGGGGTG 240

AGCTGCGTGG TCCCTTCTTT ATCACTCCTC CATCACCCCC TTCCAGTAGA GGAGTCCTTG 300

ACTTGCCAGA TATTTGGTTT TGTAGTATCA GTTCTGAAGA GCGTCTCCAT GGCTTCTCTG 360

20 GCCTGTATCA GCATTGATAG ATACATTGCC ATTACTAAAC CTTTAACCTA TAATACTCTG 420

GTTACACCCT GGAGACTACG CCTGTGTATT TTCCTGATTT GGCTATACTC GACCCTGGTC 480

TTCTGCCTT CCTTTTCCA CTGGGGCAAA CCTGGATATC ATGGAGATGT GTTTCAGTGG 540

TGTGCGGAGT CCTGGCACAC CGACTCCTAC TTCACCCTGT TCATCGTGAT GATGTTATAT 600

GCCCCAGCAG CCCTTATTGT CTGCTTCACC TATTTCAACA TCTTCCGCAT CTGCCAACAG 660

25 CACACAAAGG ATATCAGCGA AAGGCAAGCC CGCTTCAGCA GCCAGAGTGG GGAGACTGGG 720

GAAGTGCAGG CCTGTCCTGA TAAGCGCTAT AAAATGGTCC TGTTTCGAAT CACTAGTGTA 780

TTTTACATCC TCTGGTTGCC ATATATCATC TACTTCTTGT TGGAAAGCTC CACTGGCCAC 840

AGCAACCGCT TCGCATCCTT CTTGACCACC TGGCTTGCTA TTAGTAACAG TTTCTGCAAC 900

TGTGTAATTT ATAGTCTCTC CAACAGTGTA TTCCAAAGAG GACTAAAGCG CCTCTCAGGG 960

30 GCTATGTGTA CTTCTTGTGC AAGTCAGACT ACAGCCAACG ACCCTTACAC AGTTAGAAGC 1020

AAAGGCCCTC TTAATGGATG TCATATCTGA 1050

(189) INFORMATION FOR SEQ ID NO:188:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 349 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:188:

10 Met Asn Ser Thr Leu Asp Gly Asn Gln Ser Ser His Pro Phe Cys Leu
 1 5 10 15
 Leu Ala Phe Gly Tyr Leu Glu Thr Val Asn Phe Cys Leu Leu Glu Val
 20 25 30
 Leu Ile Ile Val Phe Leu Thr Val Leu Ile Ile Ser Gly Asn Ile Ile
 35 40 45
 15 Val Ile Phe Val Phe His Cys Ala Pro Leu Leu Asn His His Thr Thr
 50 55 60
 Ser Tyr Phe Ile Gln Thr Met Ala Tyr Ala Asp Leu Phe Val Gly Val
 65 70 75 80
 20 Ser Cys Val Val Pro Ser Leu Ser Leu Leu His His Pro Leu Pro Val
 85 90 95
 Glu Glu Ser Leu Thr Cys Gln Ile Phe Gly Phe Val Val Ser Val Leu
 100 105 110
 Lys Ser Val Ser Met Ala Ser Leu Ala Cys Ile Ser Ile Asp Arg Tyr
 115 120 125
 25 Ile Ala Ile Thr Lys Pro Leu Thr Tyr Asn Thr Leu Val Thr Pro Trp
 130 135 140
 Arg Leu Arg Leu Cys Ile Phe Leu Ile Trp Leu Tyr Ser Thr Leu Val
 145 150 155 160
 30 Phe Leu Pro Ser Phe Phe His Trp Gly Lys Pro Gly Tyr His Gly Asp
 165 170 175
 Val Phe Gln Trp Cys Ala Glu Ser Trp His Thr Asp Ser Tyr Phe Thr
 180 185 190
 Leu Phe Ile Val Met Met Leu Tyr Ala Pro Ala Ala Leu Ile Val Cys
 195 200 205
 35 Phe Thr Tyr Phe Asn Ile Phe Arg Ile Cys Gln Gln His Thr Lys Asp
 210 215 220

141

Ile Ser Glu Arg Gln Ala Arg Phe Ser Ser Gln Ser Gly Glu Thr Gly
 225 230 235 240

Glu Val Gln Ala Cys Pro Asp Lys Arg Tyr Lys Met Val Leu Phe Arg
 245 250 255

5 Ile Thr Ser Val Phe Tyr Ile Leu Trp Leu Pro Tyr Ile Ile Tyr Phe
 260 265 270

Leu Leu Glu Ser Ser Thr Gly His Ser Asn Arg Phe Ala Ser Phe Leu
 275 280 285

10 Thr Thr Trp Leu Ala Ile Ser Asn Ser Phe Cys Asn Cys Val Ile Tyr
 290 295 300

Ser Leu Ser Asn Ser Val Phe Gln Arg Gly Leu Lys Arg Leu Ser Gly
 305 310 315 320

Ala Met Cys Thr Ser Cys Ala Ser Gln Thr Thr Ala Asn Asp Pro Tyr
 325 330 335

15 Thr Val Arg Ser Lys Gly Pro Leu Asn Gly Cys His Ile
 340 345

(190) INFORMATION FOR SEQ ID NO:189:

(i) SEQUENCE CHARACTERISTICS:

20 (A) LENGTH: 1302 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:189:

25 ATGTGTTTTT CTCCCATTCT GGAAATCAAC ATGCAGTCTG AATCTAACAT TACAGTGCGA 60

GATGACATTG ATGACATCAA CACCAATATG TACCAACCAC TATCATATCC GTTAAGCTTT 120

CAAGTGCTCTC TCACCGGATT TCTTATGTGA GAAATTGTGT TGGGACTTGG CAGCAACCTC 180

ACTGTATTGG TACTTTACTG CATGAAATCC AACTTAATCA ACTCTGTCAG TAACATTATT 240

ACAATGAATC TTCATGTACT TGATGTAATA ATTTGTGTGG GATGTATTCC TCTAACTATA 300

30 GTTATCCTTC TGCTTTCAC TGGAGAGTAAC ACTGCTCTCA TTTGCTGTTT CCATGAGGCT 360

TGTGTATCTT TTGCAAGTGT CTCAACAGCA ATCAACGTTT TTGCTATCAC TTTGGACAGA 420

TATGACATCT CTGTAAAACC TGCAAACCGA ATTCTGACAA TGGGCAGAGC TGTAATGTGA 480

ATGATATCCA TTTGGATTTT TTCTTTTTTC TCTTTCCTGA TTCCTTTTAT TGAGGTAAAT 540

TTTTTCAGTC TTCAAAGTGG AAATACCTGG GAAAACAAGA CACTTTTATG TGTCAGTACA 600

AATGAATACT ACACTGAACT GCGAATGTAT TATCACCTGT TAGTACAGAT CCCAATATTC 660
 TTTTTCAC TG TAGTAAT GTTAATCACA TACACCAAAA TACTTCAGGC TCTTAATATT 720
 CGAATAGGCA CAAGATTTTC AACAGGGCAG AAGAAGAAAG CAAGAAAGAA AAAGACAATT 780
 TCTCTAACCA CACAACATGA GGCTACAGAC ATGTCACAAA GCAGTGGTGG GAGAAATGTA 840
 5 GTCTTTGGTG TAAGAACTTC AGTTTCTGTA ATAATTGCCC TCCGGCGAGC TGTGAAACGA 900
 CACCGTGAAC GACGAGAAAG ACAAAGAGA GTCAAGAGGA TGTCTTTATT GATTATTTCT 960
 ACATTTCTTC TCTGCTGGAC ACCAATTTCT GTTTTAAATA CCACCATTTT ATGTTTAGGC 1020
 CCAAGTGACC TTTTAGTAAA ATTAAGATTG TGTTTTTTAG TCATGGCTTA TGGAACAAC 1080
 ATATTTTACC CTCTATTATA TGCATTCAC AGACAAAAAT TTCAAAGGT CTTGAAAAGT 1140
 10 AAAATGAAAA AGCGAGTTGT TTCTATAGTA GAAGCTGATC CCCTGCCTAA TAATGCTGTA 1200
 ATACACAACT CTTGGATAGA TCCCAAAGA AACAAAAAA TTACCTTTGA AGATAGTGAA 1260
 ATAAGAGAAA AACGTTTAGT GCCTCAGGT GTACAGACT AG 1302

(191) INFORMATION FOR SEQ ID NO:190:

15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 433 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant
 (ii) MOLECULE TYPE: protein
 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:190:
 Met Cys Phe Ser Pro Ile Leu Glu Ile Asn Met Gln Ser Glu Ser Asn
 1 5 10 15
 Ile Thr Val Arg Asp Asp Ile Asp Asp Ile Asn Thr Asn Met Tyr Gln
 20 25 30
 25 Pro Leu Ser Tyr Pro Leu Ser Phe Gln Val Ser Leu Thr Gly Phe Leu
 35 40 45
 Met Leu Glu Ile Val Leu Gly Leu Gly Ser Asn Leu Thr Val Leu Val
 50 55 60
 30 Leu Tyr Cys Met Lys Ser Asn Leu Ile Asn Ser Val Ser Asn Ile Ile
 65 70 75 80
 Thr Met Asn Leu His Val Leu Asp Val Ile Ile Cys Val Gly Cys Ile
 85 90 95
 Pro Leu Thr Ile Val Ile Leu Leu Leu Ser Leu Glu Ser Asn Thr Ala

[illegible]

Ile His Asn Ser Trp Ile Asp Pro Lys Arg Asn Lys Lys Ile Thr Phe
 405 410 415

Glu Asp Ser Glu Ile Arg Glu Lys Arg Leu Val Pro Gln Val Val Thr
 420 425 430

5 Asp

(192) INFORMATION FOR SEQ ID NO:191:

(i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 1209 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:191:

15 ATGTTGTGTC CTTCCAAGAC AGATGGCTCA GGGCACTCTG GTAGGATTCA CCAGGAAACT 60
 CATGGAGAAG GGAAAAGGGA CAAGATTAGC AACAGTGAAG GGAGGGAGAA TGGTGGGAGA 120
 GGATTCCAGA TGAACGGTGG GTCGCTGGAG GCTGAGCATG CCAGCAGGAT GTCAGTTCTC 180
 AGAGCAAAGC CCATGTCAAA CAGCCAACGC TTGCTCCTTC TGTCCCCAGG ATCACCTCCT 240
 CGCACGGGGA GCATCTCCTA CATCAACATC ATCATGCCTT CGGTGTTCCG CACCATCTGC 300
 20 CTCCTGGGCA TCATCGGGAA CTCCACGGTC ATCTTCGGCG TCGTGAAGAA GTCCAAGCTG 360
 CACTGGTGCA ACAACGTCCC CGACATCTTC ATCATCAACC TCTCGGTAGT AGATCTCCTC 420
 TTTCTCCTGG GCATGCCCTT CATGATCCAC CAGCTCATGG GCAATGGGGT GTGGCACTTT 480
 GGGGAGACCA TGTGCACCTT CATCACGGCC ATGGATGCCA ATAGTCAGTT CACCAGCACC 540
 TACATCCTGA CCGCCATGGC CATTGACCGC TACCTGGCCA CTGTCCACCC CATCTCTTCC 600
 25 ACGAAGTTCC GGAAGCCCTC TGTGGCCACC CTGGTGATCT GCCTCCTGTG GGCCCTCTCC 660
 TTCATCAGCA TCACCCTGT GTGGCTGTAT GCCAGACTCA TCCCCTTCCC AGGAGGTGCA 720
 GTGGGCTGCG GCATACGCCT GCCCAACCCA GACACTGACC TCTACTGGTT CACCCTGTAC 780
 CAGTTTTTCC TGGCCTTTGC CCTGCCTTTT GTGGTCATCA CAGCCGCATA CGTGAGGATC 840
 CTGCAGCGCA TGACGTCTC AGTGGCCCCC GCCTCCAGC GCAGCATCCG GCTGCGGACA 900
 30 AAGAGGGTGA AACGCACAGC CATCGCCATC TGTCTGGTCT TCTTTGTGTG CTGGGCACCC 960
 TACTATGTGC TACAGCTGAC CCAGTTGTCC ATCAGCCGCC CGACCCTCAC CTTTGTCTAC 1020
 TTATACAATG CGGCCATCAG CTTGGGCTAT GCCAACAGCT GCCTCAACCC CTTTGTGTAC 1080

145

ATCGTGCTCT GTGAGACGTT CCGCAAACGC TTGGTCCTGT CGGTGAAGCC TGCAGCCCAG 1140
 GGGCAGCTTC GCGCTGTCAG CAACGCTCAG ACGGCTGACG AGGAGAGGAC AGAAAGCAAA 1200
 GGCACCTCA 1209

(193) INFORMATION FOR SEQ ID NO:192:

- 5 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 402 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

- 10 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:192:

	Met	Leu	Cys	Pro	Ser	Lys	Thr	Asp	Gly	Ser	Gly	His	Ser	Gly	Arg	Ile	
	1				5					10					15		
	His	Gln	Glu	Thr	His	Gly	Glu	Gly	Lys	Arg	Asp	Lys	Ile	Ser	Asn	Ser	
15				20					25					30			
	Glu	Gly	Arg	Glu	Asn	Gly	Gly	Arg	Gly	Phe	Gln	Met	Asn	Gly	Gly	Ser	
			35					40					45				
	Leu	Glu	Ala	Glu	His	Ala	Ser	Arg	Met	Ser	Val	Leu	Arg	Ala	Lys	Pro	
		50					55					60					
20	Met	Ser	Asn	Ser	Gln	Arg	Leu	Leu	Leu	Leu	Ser	Pro	Gly	Ser	Pro	Pro	
	65					70					75					80	
	Arg	Thr	Gly	Ser	Ile	Ser	Tyr	Ile	Asn	Ile	Ile	Met	Pro	Ser	Val	Phe	
					85					90					95		
	Gly	Thr	Ile	Cys	Leu	Leu	Gly	Ile	Ile	Gly	Asn	Ser	Thr	Val	Ile	Phe	
25				100					105					110			
	Ala	Val	Val	Lys	Lys	Ser	Lys	Leu	His	Trp	Cys	Asn	Asn	Val	Pro	Asp	
			115					120					125				
	Ile	Phe	Ile	Ile	Asn	Leu	Ser	Val	Val	Asp	Leu	Leu	Phe	Leu	Leu	Gly	
		130				135						140					
30	Met	Pro	Phe	Met	Ile	His	Gln	Leu	Met	Gly	Asn	Gly	Val	Trp	His	Phe	
	145					150					155					160	
	Gly	Glu	Thr	Met	Cys	Thr	Leu	Ile	Thr	Ala	Met	Asp	Ala	Asn	Ser	Gln	
				165						170					175		
	Phe	Thr	Ser	Thr	Tyr	Ile	Leu	Thr	Ala	Met	Ala	Ile	Asp	Arg	Tyr	Leu	
35				180					185					190			
	Ala	Thr	Val	His	Pro	Ile	Ser	Ser	Thr	Lys	Phe	Arg	Lys	Pro	Ser	Val	

195 200 205

Ala Thr Leu Val Ile Cys Leu Leu Trp Ala Leu Ser Phe Ile Ser Ile
210 215 220

5 Thr Pro Val Trp Leu Tyr Ala Arg Leu Ile Pro Phe Pro Gly Gly Ala
225 230 235 240

Val Gly Cys Gly Ile Arg Leu Pro Asn Pro Asp Thr Asp Leu Tyr Trp
245 250 255

Phe Thr Leu Tyr Gln Phe Phe Leu Ala Phe Ala Leu Pro Phe Val Val
260 265 270

10 Ile Thr Ala Ala Tyr Val Arg Ile Leu Gln Arg Met Thr Ser Ser Val
275 280 285

Ala Pro Ala Ser Gln Arg Ser Ile Arg Leu Arg Thr Lys Arg Val Lys
290 295 300

15 Arg Thr Ala Ile Ala Ile Cys Leu Val Phe Phe Val Cys Trp Ala Pro
305 310 315 320

Tyr Tyr Val Leu Gln Leu Thr Gln Leu Ser Ile Ser Arg Pro Thr Leu
325 330 335

Thr Phe Val Tyr Leu Tyr Asn Ala Ala Ile Ser Leu Gly Tyr Ala Asn
340 345 350

20 Ser Cys Leu Asn Pro Phe Val Tyr Ile Val Leu Cys Glu Thr Phe Arg
355 360 365

Lys Arg Leu Val Leu Ser Val Lys Pro Ala Ala Gln Gly Gln Leu Arg
370 375 380

25 Ala Val Ser Asn Ala Gln Thr Ala Asp Glu Glu Arg Thr Glu Ser Lys
385 390 395 400

Gly Thr

(194) INFORMATION FOR SEQ ID NO:193:

- 30 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1128 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- 35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:193:

ATGGATGTGA CTTCCAAGC CCGGGGCGTG GGCCTGGAGA TGTACCCAGG CACCGCGCAC 60

GCTGCGGCCC CCAACACCAC CTCCCCGAG CTCAACCTGT CCCACCGCT CCTGGGCACC 120

147

GCCCTGGCCA ATGGGACAGG TGAGCTCTCG GAGCACCAGC AGTACGTGAT CGGCCTGTTC 180
 CTCTCGTGCC TCTACACCAT CTTCTCTTTC CCCATCGGCT TTGTGGGCAA CATCTGATC 240
 CTGGTGGTGA ACATCAGCTT CCGCGAGAAG ATGACCATCC CCGACCTGTA CTTCATCAAC 300
 CTGGCGGTGG CGGACCTCAT CCTGGTGGCC GACTCCCTCA TTGAGGTGTT CAACCTGCAC 360
 5 GAGCGGTACT ACGACATCGC CGTCCTGTGC ACCTTCATGT CGCTCTTCCT GCAGGTCAAC 420
 ATGTACAGCA GCGTCTTCTT CCTCACCTGG ATGAGCTTCG ACCGCTACAT CGCCCTGGCC 480
 AGGGCCATGC GCTGCAGCCT GTTCCGCACC AAGCACCACG CCCGGCTGAG CTGTGGCCTC 540
 ATCTGGATGG CATCCGTGTC AGCCACGCTG GTGCCCTTCA CCGCCGTGCA CCTGCAGCAC 600
 ACCGACGAGG CTGCTTCTG TTTGCGGGAT GTCCGGGAGG TGCAGTGGCT CGAGGTCACG 660
 10 CTGGGCTTCA TCGTGCCCTT CGCCATCATC GGCCTGTGCT ACTCCCTCAT TGTCCGGGTG 720
 CTGGTCAGGG CGCACC GGCA CCGTGGGCTG CGGCCCCGGC GGCAGAAGGC GAAACGCATG 780
 ATCCTCGCGG TGGTGCTGGT CTTCTTCGTC TGCTGGCTGC CGGAGAACGT CTTCATCAGC 840
 GTGCACCTCC TGCAGCGGAC GCAGCCTGGG GCCGCTCCCT GCAAGCAGTC TTTCCGCCAT 900
 GCCCACCCTC TCACGGGCCA CATTGTCAAC CTCGCCGCCT TCTCCAACAG CTGCCTAAAC 960
 15 CCCCTCATCT ACAGCTTTCT CGGGGAGACC TTCAGGGACA AGCTGAGGCT GTACATTGAG 1020
 CAGAAAACAA ATTTGCCGGC CCTGAACCGC TTCTGTCACG CTGCCCTGAA GGCCGTCATT 1080
 CCAGACAGCA CCGAGCAGTC GGATGTGAGG TTCAGCAGTG CCGTGTGA 1128

(195) INFORMATION FOR SEQ ID NO:194:

- 20 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 375 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

- 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:194:

Met Asp Val Thr Ser Gln Ala Arg Gly Val Gly Leu Glu Met Tyr Pro
 1 5 10 15

Gly Thr Ala His Ala Ala Ala Pro Asn Thr Thr Ser Pro Glu Leu Asn
 20 25 30

- 30 Leu Ser His Pro Leu Leu Gly Thr Ala Leu Ala Asn Gly Thr Gly Glu
 35 40 45

Leu Ser Glu His Gln Gln Tyr Val Ile Gly Leu Phe Leu Ser Cys Leu
 50 55 60
 Tyr Thr Ile Phe Leu Phe Pro Ile Gly Phe Val Gly Asn Ile Leu Ile
 65 70 75 80
 5 Leu Val Val Asn Ile Ser Phe Arg Glu Lys Met Thr Ile Pro Asp Leu
 85 90 95
 Tyr Phe Ile Asn Leu Ala Val Ala Asp Leu Ile Leu Val Ala Asp Ser
 100 105 110
 10 Leu Ile Glu Val Phe Asn Leu His Glu Arg Tyr Tyr Asp Ile Ala Val
 115 120 125
 Leu Cys Thr Phe Met Ser Leu Phe Leu Gln Val Asn Met Tyr Ser Ser
 130 135 140
 Val Phe Phe Leu Thr Trp Met Ser Phe Asp Arg Tyr Ile Ala Leu Ala
 145 150 155 160
 15 Arg Ala Met Arg Cys Ser Leu Phe Arg Thr Lys His His Ala Arg Leu
 165 170 175
 Ser Cys Gly Leu Ile Trp Met Ala Ser Val Ser Ala Thr Leu Val Pro
 180 185 190
 20 Phe Thr Ala Val His Leu Gln His Thr Asp Glu Ala Cys Phe Cys Phe
 195 200 205
 Ala Asp Val Arg Glu Val Gln Trp Leu Glu Val Thr Leu Gly Phe Ile
 210 215 220
 Val Pro Phe Ala Ile Ile Gly Leu Cys Tyr Ser Leu Ile Val Arg Val
 225 230 235 240
 25 Leu Val Arg Ala His Arg His Arg Gly Leu Arg Pro Arg Arg Gln Lys
 245 250 255
 Ala Lys Arg Met Ile Leu Ala Val Val Leu Val Phe Phe Val Cys Trp
 260 265 270
 30 Leu Pro Glu Asn Val Phe Ile Ser Val His Leu Leu Gln Arg Thr Gln
 275 280 285
 Pro Gly Ala Ala Pro Cys Lys Gln Ser Phe Arg His Ala His Pro Leu
 290 295 300
 Thr Gly His Ile Val Asn Leu Ala Ala Phe Ser Asn Ser Cys Leu Asn
 305 310 315 320
 35 Pro Leu Ile Tyr Ser Phe Leu Gly Glu Thr Phe Arg Asp Lys Leu Arg
 325 330 335
 Leu Tyr Ile Glu Gln Lys Thr Asn Leu Pro Ala Leu Asn Arg Phe Cys

149

340 345 350

His Ala Ala Leu Lys Ala Val Ile Pro Asp Ser Thr Glu Gln Ser Asp
355 360 365

Val Arg Phe Ser Ser Ala Val
370 375

5

(196) INFORMATION FOR SEQ ID NO:195:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 960 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:195:

ATGCCATTCC CAAACTGCTC AGCCCCCAGC ACTGTGGTGG CCACAGCTGT GGGTGTCTTG 60

15 CTGGGGCTGG AGTGTGGGCT GGGTCTGCTG GGCAACGCGG TGGCGCTGTG GACCTTCCTG 120

TTCCGGGTCA GGGTGTGGAA GCCGTACGCT GTCTACCTGC TCAACCTGGC CCTGGCTGAC 180

CTGCTGTTGG CTGCGTGCCT GCCTTTCCTG GCCGCCTTCT ACCTGAGCCT CCAGGCTTGG 240

CATCTGGGCC GTGTGGGCTG CTGGGCCCTG CGCTTCCTGC TGGACCTCAG CCGCAGCGTG 300

GGGATGGCCT TCCTGGCCGC CGTGGCTTTG GACCGGTACC TCCGTGTGGT CCACCCTCGG 360

20 CTTAAGGTCA ACCTGCTGTC TCCTCAGGCG GCCCTGGGGG TCTCGGGCCT CGTCTGGCTC 420

CTGATGGTCG CCCTCACCTG CCCGGGCTTG CTCATCTCTG AGGCCGCCCA GAACTCCACC 480

AGGTGCCACA GTTTCTACTC CAGGGCAGAC GGCTCCTTCA GCATCATCTG GCAGGAAGCA 540

CTCTCCTGCC TTCAGTTTGT CCTCCCCTTT GGCCTCATCG TGTTCGTCAA TGCAGGCATC 600

ATCAGGGCTC TCCAGAAAAG ACTCCGGGAG CCTGAGAAAC AGCCCAAGCT TCAGCGGGCC 660

25 AAGGCACTGG TCACCTTGGT GGTGGTGCTG TTTGCTCTGT GCTTTCTGCC CTGCTTCCTG 720

GCCAGAGTCC TGATGCACAT CTTCCAGAAT CTGGGGAGCT GCAGGGCCCT TTGTGCAGTG 780

GCTCATACCT CGGATGTCAC GGGCAGCCTC ACCTACCTGC ACAGTGTCTG CAACCCCGTG 840

GTATACTGCT TCTCCAGCCC CACCTTCAGG AGCTCCTATC GGAGGGTCTT CCACACCCTC 900

CGAGGCAAAG GGCAGGCAGC AGAGCCCCCA GATTTC AACC CCAGAGACTC CTATTCCTGA 960

30 (197) INFORMATION FOR SEQ ID NO:196:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 319 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

5 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:196:

```

Met Pro Phe Pro Asn Cys Ser Ala Pro Ser Thr Val Val Ala Thr Ala
1          5          10          15
Val Gly Val Leu Leu Gly Leu Glu Cys Gly Leu Gly Leu Leu Gly Asn
10          20          25          30
Ala Val Ala Leu Trp Thr Phe Leu Phe Arg Val Arg Val Trp Lys Pro
          35          40          45
Tyr Ala Val Tyr Leu Leu Asn Leu Ala Leu Ala Asp Leu Leu Leu Ala
          50          55          60
15 Ala Cys Leu Pro Phe Leu Ala Ala Phe Tyr Leu Ser Leu Gln Ala Trp
65          70          75          80
His Leu Gly Arg Val Gly Cys Trp Ala Leu Arg Phe Leu Leu Asp Leu
          85          90          95
20 Ser Arg Ser Val Gly Met Ala Phe Leu Ala Ala Val Ala Leu Asp Arg
          100          105          110
Tyr Leu Arg Val Val His Pro Arg Leu Lys Val Asn Leu Leu Ser Pro
          115          120          125
Gln Ala Ala Leu Gly Val Ser Gly Leu Val Trp Leu Leu Met Val Ala
          130          135          140
25 Leu Thr Cys Pro Gly Leu Leu Ile Ser Glu Ala Ala Gln Asn Ser Thr
145          150          155          160
Arg Cys His Ser Phe Tyr Ser Arg Ala Asp Gly Ser Phe Ser Ile Ile
          165          170          175
30 Trp Gln Glu Ala Leu Ser Cys Leu Gln Phe Val Leu Pro Phe Gly Leu
          180          185          190
Ile Val Phe Cys Asn Ala Gly Ile Ile Arg Ala Leu Gln Lys Arg Leu
          195          200          205
Arg Glu Pro Glu Lys Gln Pro Lys Leu Gln Arg Ala Lys Ala Leu Val
          210          215          220
35 Thr Leu Val Val Val Leu Phe Ala Leu Cys Phe Leu Pro Cys Phe Leu
225          230          235          240
Ala Arg Val Leu Met His Ile Phe Gln Asn Leu Gly Ser Cys Arg Ala

```

151

				245					250						255				
Leu	Cys	Ala	Val	Ala	His	Thr	Ser	Asp	Val	Thr	Gly	Ser	Leu	Thr	Tyr				
			260					265						270					
Leu	His	Ser	Val	Val	Asn	Pro	Val	Val	Tyr	Cys	Phe	Ser	Ser	Pro	Thr				
		275					280					285							
Phe	Arg	Ser	Ser	Tyr	Arg	Arg	Val	Phe	His	Thr	Leu	Arg	Gly	Lys	Gly				
	290					295					300								
Gln	Ala	Ala	Glu	Pro	Pro	Asp	Phe	Asn	Pro	Arg	Asp	Ser	Tyr	Ser					
305					310					315									

10 (198) INFORMATION FOR SEQ ID NO:197:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1143 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:197:

	ATGGAGGAAG	GTGGTGATTT	TGACAACTAC	TATGGGGGCAG	ACAACCAGTC	TGAGTGTGAG	60
	TACACAGACT	GGAAATCCTC	GGGGGCCCTC	ATCCCTGCCA	TCTACATGTT	GGTCTTCCTC	120
20	CTGGGCACCA	CGGGAAACGG	TCTGGTGCTC	TGGACCGTGT	TTCGGAGCAG	CCGGGAGAAG	180
	AGGCGCTCAG	CTGATATCTT	CATTGCTAGC	CTGGCGGTGG	CTGACCTGAC	CTTCGTGGTG	240
	ACGCTGCCCC	TGTGGGCTAC	CTACACGTAC	CGGGACTATG	ACTGGCCCTT	TGGGACCTTC	300
	TTCTGCAAGC	TCAGCAGCTA	CCTCATCTTC	GTCAACATGT	ACGCCAGCGT	CTTCTGCCTC	360
	ACCGGCCTCA	GCTTCGACCG	CTACCTGGCC	ATCGTGAGGC	CAGTGGCCAA	TGCTCGGCTG	420
25	AGGCTGCGGG	TCAGCGGGGC	CGTGGCCACG	GCAGTTCTTT	GGGTGCTGGC	CGCCCTCCTG	480
	GCCATGCCTG	TCATGGTGTT	ACGCACCACC	GGGGACTTGG	AGAACACCAC	TAAGGTGCAG	540
	TGCTACATGG	ACTACTCCAT	GGTGGCCACT	GTGAGCTCAG	AGTGGGCCTG	GGAGGTGGGC	600
	CTTGGGGTCT	CGTCCACCAC	CGTGGGCTTT	GTGGTGCCCT	TCACCATCAT	GCTGACCTGT	660
	TACTTCTTCA	TCGCCCCAAC	CATCGCTGGC	CACTTCCGCA	AGGAACGCAT	CGAGGGCCTG	720
30	CGGAAGCGGC	GCCGGCTTAA	GAGCATCATC	GTGGTGCTGG	TGGTGACCTT	TGCCCTGTGC	780
	TGGATGCCCT	ACCACCTGGT	GAAGACGCTG	TACATGCTGG	GCAGCCTGCT	GCACTGGCCC	840
	TGTGACTTTG	ACCTCTTCCT	CATGAACATC	TTCCCCTACT	GCACCTGCAT	CAGCTACGTC	900

AACAGCTGCC TCAACCCCTT CCTCTATGCC TTTTTCGACC CCCGCTTCCG CCAGGCCTGC 960
 ACCTCCATGC TCTGCTGTGG CCAGAGCAGG TGCGCAGGCA CCTCCCACAG CAGCAGTGGG 1020
 GAGAAGTCAG CCAGCTACTC TTCGGGGCAC AGCCAGGGGC CCGGCCCAA CATCGGCAAG 1080
 GGTGGAGAAC AGATGCACGA GAAATCCATC CCCTACAGCC AGGAGACCCT TGTGGTTGAC 1140
 5 TAG 1143

(199) INFORMATION FOR SEQ ID NO:198:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 380 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:198:

15 Met Glu Glu Gly Gly Asp Phe Asp Asn Tyr Tyr Gly Ala Asp Asn Gln
 1 5 10 15
 Ser Glu Cys Glu Tyr Thr Asp Trp Lys Ser Ser Gly Ala Leu Ile Pro
 20 25 30
 Ala Ile Tyr Met Leu Val Phe Leu Leu Gly Thr Thr Gly Asn Gly Leu
 35 40 45
 20 Val Leu Trp Thr Val Phe Arg Ser Ser Arg Glu Lys Arg Arg Ser Ala
 50 55 60
 Asp Ile Phe Ile Ala Ser Leu Ala Val Ala Asp Leu Thr Phe Val Val
 65 70 75 80
 25 Thr Leu Pro Leu Trp Ala Thr Tyr Thr Tyr Arg Asp Tyr Asp Trp Pro
 85 90 95
 Phe Gly Thr Phe Phe Cys Lys Leu Ser Ser Tyr Leu Ile Phe Val Asn
 100 105 110
 Met Tyr Ala Ser Val Phe Cys Leu Thr Gly Leu Ser Phe Asp Arg Tyr
 115 120 125
 30 Leu Ala Ile Val Arg Pro Val Ala Asn Ala Arg Leu Arg Leu Arg Val
 130 135 140
 Ser Gly Ala Val Ala Thr Ala Val Leu Trp Val Leu Ala Ala Leu Leu
 145 150 155 160
 35 Ala Met Pro Val Met Val Leu Arg Thr Thr Gly Asp Leu Glu Asn Thr
 165 170 175

153

Thr Lys Val Gln Cys Tyr Met Asp Tyr Ser Met Val Ala Thr Val Ser
 180 185 190
 Ser Glu Trp Ala Trp Glu Val Gly Leu Gly Val Ser Ser Thr Thr Val
 195 200 205
 5 Gly Phe Val Val Pro Phe Thr Ile Met Leu Thr Cys Tyr Phe Phe Ile
 210 215 220
 Ala Gln Thr Ile Ala Gly His Phe Arg Lys Glu Arg Ile Glu Gly Leu
 225 230 235 240
 10 Arg Lys Arg Arg Arg Leu Lys Ser Ile Ile Val Val Leu Val Val Thr
 245 250 255
 Phe Ala Leu Cys Trp Met Pro Tyr His Leu Val Lys Thr Leu Tyr Met
 260 265 270
 Leu Gly Ser Leu Leu His Trp Pro Cys Asp Phe Asp Leu Phe Leu Met
 275 280 285
 15 Asn Ile Phe Pro Tyr Cys Thr Cys Ile Ser Tyr Val Asn Ser Cys Leu
 290 295 300
 Asn Pro Phe Leu Tyr Ala Phe Phe Asp Pro Arg Phe Arg Gln Ala Cys
 305 310 315 320
 20 Thr Ser Met Leu Cys Cys Gly Gln Ser Arg Cys Ala Gly Thr Ser His
 325 330 335
 Ser Ser Ser Gly Glu Lys Ser Ala Ser Tyr Ser Ser Gly His Ser Gln
 340 345 350
 Gly Pro Gly Pro Asn Met Gly Lys Gly Gly Glu Gln Met His Glu Lys
 355 360 365
 25 Ser Ile Pro Tyr Ser Gln Glu Thr Leu Val Val Asp
 370 375 380

(200) INFORMATION FOR SEQ ID NO:199:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1119 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: DNA (genomic)
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:199:
 35 ATGAACTACC CGCTAACGCT GGAAATGGAC CTCGAGAACC TGGAGGACCT GTTCTGGGAA 60
 CTGGACAGAT TGGACAAC TAACGACACC TCCCTGGTGG AAAATCATCT CTGCCCTGCC 120

ACAGAGGGTC CCCTCATGGC CTCCTTCAAG GCCGTGTTTCG TGCCCGTGGC CTACAGCCTC 180
 ATCTTCCTCC TGGGCGTGAT CGGCAACGTC CTGGTGCTGG TGATCCTGGA GCGGCACCGG 240
 CAGACACGCA GTTCCACGGA GACCTTCCTG TTCCACCTGG CCGTGGCCGA CCTCCTGCTG 300
 GTCTTCATCT TGCCCTTTGC CGTGGCCGAG GGCTCTGTGG GCTGGGTCTT GGGGACCTTC 360
 5 CTCTGCAAAA CTGTGATTGC CCTGCACAAA GTCAACTTCT ACTGCAGCAG CCTGCTCCTG 420
 GCCTGCATCG CCGTGGACCG CTACCTGGCC ATTGTCCACG CCGTCCATGC CTACCGCCAC 480
 CGCCGCCTCC TCTCCATCCA CATCACCTGT GGGACCATCT GGCTGGTGGG CTTCCTCCTT 540
 GCCTTGCCAG AGATTCTCTT CGCCAAAGTC AGCCAAGGCC ATCACAACAA CTCCCTGCCA 600
 CGTTGCACCT TCTCCCAAGA GAACCAAGCA GAAACGCATG CCTGGTTCAC CTCCCGATTC 660
 10 CTCTACCATG TGGCGGGATT CCTGCTGCCC ATGCTGGTGA TGGGCTGGTG CTACGTGGGG 720
 GTAGTGCACA GGTTCGCGCA GGCCAGCGG CGCCCTCAGC GGCAGAAGGC AAAAAGGGTG 780
 GCCATCCTGG TGACAAGCAT CTTCTTCCTC TGCTGGTCAC CCTACCACAT CGTCATCTTC 840
 CTGGACACCC TGGCGAGGCT GAAGGCCGTG GACAATACCT GCAAGCTGAA TGGCTCTCTC 900
 CCCGTGGCCA TCACCATGTG TGAGTTCCTG GGCCTGGCCC ACTGCTGCCT CAACCCCATG 960
 15 CTCTACACTT TCGCCGGCGT GAAGTTCGC AGTGACCTGT CGCGGCTCCT GACCAAGCTG 1020
 GGCTGTACCG GCCCTGCCTC CCTGTGCCAG CTCTTCCTA GCTGGCGCAG GAGCAGTCTC 1080
 TCTGAGTCAG AGAATGCCAC CTCTCTCACC ACGTTCTAG 1119

(201) INFORMATION FOR SEQ ID NO:200:

- 20 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 372 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant
 (ii) MOLECULE TYPE: protein

- 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:200:

Met Asn Tyr Pro Leu Thr Leu Glu Met Asp Leu Glu Asn Leu Glu Asp
 1 5 10 15

Leu Phe Trp Glu Leu Asp Arg Leu Asp Asn Tyr Asn Asp Thr Ser Leu
 20 25 30

- 30 Val Glu Asn His Leu Cys Pro Ala Thr Glu Gly Pro Leu Met Ala Ser
 35 40 45

Phe Lys Ala Val Phe Val Pro Val Ala Tyr Ser Leu Ile Phe Leu Leu

155

	50		55		60															
	Gly	Val	Ile	Gly	Asn	Val	Leu	Val	Leu	Val	Ile	Leu	Glu	Arg	His	Arg				
	65				70						75					80				
5	Gln	Thr	Arg	Ser	Ser	Thr	Glu	Thr	Phe	Leu	Phe	His	Leu	Ala	Val	Ala				
				85						90					95					
	Asp	Leu	Leu	Leu	Val	Phe	Ile	Leu	Pro	Phe	Ala	Val	Ala	Glu	Gly	Ser				
				100					105					110						
	Val	Gly	Trp	Val	Leu	Gly	Thr	Phe	Leu	Cys	Lys	Thr	Val	Ile	Ala	Leu				
			115					120					125							
10	His	Lys	Val	Asn	Phe	Tyr	Cys	Ser	Ser	Leu	Leu	Leu	Ala	Cys	Ile	Ala				
		130					135						140							
	Val	Asp	Arg	Tyr	Leu	Ala	Ile	Val	His	Ala	Val	His	Ala	Tyr	Arg	His				
	145					150					155					160				
15	Arg	Arg	Leu	Leu	Ser	Ile	His	Ile	Thr	Cys	Gly	Thr	Ile	Trp	Leu	Val				
					165					170					175					
	Gly	Phe	Leu	Leu	Ala	Leu	Pro	Glu	Ile	Leu	Phe	Ala	Lys	Val	Ser	Gln				
				180					185					190						
	Gly	His	His	Asn	Asn	Ser	Leu	Pro	Arg	Cys	Thr	Phe	Ser	Gln	Glu	Asn				
			195					200					205							
20	Gln	Ala	Glu	Thr	His	Ala	Trp	Phe	Thr	Ser	Arg	Phe	Leu	Tyr	His	Val				
		210					215					220								
	Ala	Gly	Phe	Leu	Leu	Pro	Met	Leu	Val	Met	Gly	Trp	Cys	Tyr	Val	Gly				
	225					230					235				240					
25	Val	Val	His	Arg	Leu	Arg	Gln	Ala	Gln	Arg	Arg	Pro	Gln	Arg	Gln	Lys				
					245					250					255					
	Ala	Lys	Arg	Val	Ala	Ile	Leu	Val	Thr	Ser	Ile	Phe	Phe	Leu	Cys	Trp				
				260					265					270						
	Ser	Pro	Tyr	His	Ile	Val	Ile	Phe	Leu	Asp	Thr	Leu	Ala	Arg	Leu	Lys				
			275					280					285							
30	Ala	Val	Asp	Asn	Thr	Cys	Lys	Leu	Asn	Gly	Ser	Leu	Pro	Val	Ala	Ile				
		290					295					300								
	Thr	Met	Cys	Glu	Phe	Leu	Gly	Leu	Ala	His	Cys	Cys	Leu	Asn	Pro	Met				
	305					310				315					320					
35	Leu	Tyr	Thr	Phe	Ala	Gly	Val	Lys	Phe	Arg	Ser	Asp	Leu	Ser	Arg	Leu				
				325					330						335					
	Leu	Thr	Lys	Leu	Gly	Cys	Thr	Gly	Pro	Ala	Ser	Leu	Cys	Gln	Leu	Phe				
				340					345					350						

Pro Ser Trp Arg Arg Ser Ser Leu Ser Glu Ser Glu Asn Ala Thr Ser
 355 360 365

Leu Thr Thr Phe
 370

5 (202) INFORMATION FOR SEQ ID NO:201:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1128 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 10 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:201:

ATGGATGTGA CTTCCAAGC CCGGGGCGTG GGCCTGGAGA TGTACCCAGG CACCGCGCAG 60
 CCTGCGGCCC CCAACACCAC CTCCCCGAG CTCAACCTGT CCCACCCGCT CCTGGGCACC 120
 15 GCCCTGGCCA ATGGGACAGG TGAGCTCTCG GAGCACCAGC AGTACGTGAT CGGCCTGTTC 180
 CTCTCGTGCC TCTACACCAT CTTCCTCTTC CCCATCGGCT TTGTGGGCAA CATCCTGATC 240
 CTGGTGGTGA ACATCAGCTT CCGCGAGAAG ATGACCATCC CCGACCTGTA CTTCATCAAC 300
 CTGGCGGTGG CGGACCTCAT CCTGGTGGCC GACTCCCTCA TTGAGGTGTT CAACCTGCAC 360
 GAGCGGTACT ACGACATCGC CGTCCTGTGC ACCTTCATGT CGCTCTTCCT GCAGGTCAAC 420
 20 ATGTACAGCA GCGTCTTCTT CCTCACCTGG ATGAGCTTCG ACCGCTACAT CGCCCTGGCC 480
 AGGGCCATGC GCTGCAGCCT GTTCCGCACC AAGCACCACG CCCGGCTGAG CTGTGGCCTC 540
 ATCTGGATGG CATCCGTGTC AGCCACGCTG GTGCCCTTCA CCGCCGTGCA CCTGCAGCAC 600
 ACCGACGAGG CCTGCTTCTG TTTCGCGGAT GTCCGGGAGG TGCAGTGGCT CGAGGTCACG 660
 CTGGGCTTCA TCGTGCCCTT CGCCATCATC GGCCTGTGCT ACTCCCTCAT TGTCCGGGTG 720
 25 CTGGTCAGGG CGCACCAGCA CCGTGGGCTG CGGCCCCGGC GGCAGAAGGC GAAGCGCATG 780
 ATCCTCGCGG TGGTGCTGGT CTTCTTCGTC TGCTGGCTGC CGGAGAACGT CTTATCAGC 840
 GTGCACCTCC TGCAGCGGAC GCAGCCTGGG GCCGCTCCCT GCAAGCAGTC TTTCCGCCAT 900
 GCCACCCCC TCACGGGCCA CATTGTCAAC CTCACCGCCT TCTCCAACAG CTGCCTAAAC 960
 CCCCTCATCT ACAGCTTTCT CGGGGAGACC TTCAGGGACA AGCTGAGGCT GTACATTGAG 1020
 30 CAGAAAACAA ATTTGCCGGC CCTGAACCGC TTCTGTCACG CTGCCCTGAA GGCCGTCATT 1080
 CCAGACAGCA CCGAGCAGTC GGATGTGAGG TTCAGCAGTG CCGTGTAG 1128

157

(203) INFORMATION FOR SEQ ID NO:202:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 375 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:202:

10 Met Asp Val Thr Ser Gln Ala Arg Gly Val Gly Leu Glu Met Tyr Pro
 1 5 10 15
 Gly Thr Ala Gln Pro Ala Ala Pro Asn Thr Thr Ser Pro Glu Leu Asn
 20 25 30
 Leu Ser His Pro Leu Leu Gly Thr Ala Leu Ala Asn Gly Thr Gly Glu
 35 40 45
 15 Leu Ser Glu His Gln Gln Tyr Val Ile Gly Leu Phe Leu Ser Cys Leu
 50 55 60
 Tyr Thr Ile Phe Leu Phe Pro Ile Gly Phe Val Gly Asn Ile Leu Ile
 65 70 75 80
 20 Leu Val Val Asn Ile Ser Phe Arg Glu Lys Met Thr Ile Pro Asp Leu
 85 90 95
 Tyr Phe Ile Asn Leu Ala Val Ala Asp Leu Ile Leu Val Ala Asp Ser
 100 105 110
 Leu Ile Glu Val Phe Asn Leu His Glu Arg Tyr Tyr Asp Ile Ala Val
 115 120 125
 25 Leu Cys Thr Phe Met Ser Leu Phe Leu Gln Val Asn Met Tyr Ser Ser
 130 135 140
 Val Phe Phe Leu Thr Trp Met Ser Phe Asp Arg Tyr Ile Ala Leu Ala
 145 150 155 160
 30 Arg Ala Met Arg Cys Ser Leu Phe Arg Thr Lys His His Ala Arg Leu
 165 170 175
 Ser Cys Gly Leu Ile Trp Met Ala Ser Val Ser Ala Thr Leu Val Pro
 180 185 190
 Phe Thr Ala Val His Leu Gln His Thr Asp Glu Ala Cys Phe Cys Phe
 195 200 205
 35 Ala Asp Val Arg Glu Val Gln Trp Leu Glu Val Thr Leu Gly Phe Ile
 210 215 220
 Val Pro Phe Ala Ile Ile Gly Leu Cys Tyr Ser Leu Ile Val Arg Val

20 (204) INFORMATION FOR SEQ ID NO:203:

(i) SEQUENCE CHARACTERISTICS:

```

25      (C) STRANDEDNESS: single
      (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

```

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:203:

30	TACACTTTGT TCGAGTCTTT GTGCTCCAAG AAGGACGTGC GGAAC TTAA AGCCTGGTTC	180
	CTCCCTATCA TGTACTCCAT CATTTGTTTC GTGGGCCTAC TGGGCAATGG GCTGGTCGTG	240
	TTGACCTATA TCTATTTCAA GAGGCTCAAG ACCATGACCG ATACCTACCT GCTCAACCTG	300
	GCGGTGGCAG ACATCCTCTT CCTCCTGACC CTTCCCTTCT GGGCCTACAG CGCGGCCAAG	360
	TCCTGGGTCT TCGGTGTCCA CTTTTCGAAG CTCATCTTTG CCATCTACAA GATGAGCTTC	420

159

TTCAGTGGCA TGCTCCTACT TCTTTGCATC AGCATTGACC GCTACGTGGC CATCGTCCAG 480
 GCTGTCTCAG CTCACCGCCA CCGTGCCCCG GTCCTTCTCA TCAGCAAGCT GTCCTGTGTG 540
 GGCATCTGGA TACTAGCCAC AGTGCTCTCC ATCCCAGAGC TCCTGTACAG TGACCTCCAG 600
 AGGAGCAGCA GTGAGCAAGC GATGCGATGC TCTCTCATCA CAGAGCATGT GGAGGCCTTT 660
 5 ATCACCATCC AGGTGGCCCA GATGGTGATC GGCTTTCTGG TCCCCCTGCT GGCCATGAGC 720
 TTCTGTTACC TTGTCATCAT CCGCACCCTG CTCCAGGCAC GCAACTTTGA GCGCAACAAG 780
 GCCAAAAGG TGATCATCGC TGTGGTCGTG GTCTTCATAG TCTTCCAGCT GCCCTACAAT 840
 GGGGTGGTCC TGGCCCAGAC GGTGGCCAAC TTCAACATCA CCAGTAGCAC CTGTGAGCTC 900
 AGTAAGCAAC TCAACATCGC CTACGACGTC ACCTACAGCC TGGCCTGCGT CCGCTGCTGC 960
 10 GTCAACCCTT TCTTGTACGC CTTTCATCGGC GTCAAGTTCC GCAACGATCT CTTCAAGCTC 1020
 TTCAAGGACC TGGGCTGCCT CAGCCAGGAG CAGCTCCGGC AGTGGTCTTC CTGTGGGCAC 1080
 ATCCGGCGCT CCTCCATGAG TGTGGAGGCC GAGACCACCA CCACCTTCTC CCCATAG 1137

(205) INFORMATION FOR SEQ ID NO:204:

- 15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 378 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

- 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:204:

Met Asp Leu Gly Lys Pro Met Lys Ser Val Leu Val Val Ala Leu Leu
 1 5 10 15

Val Ile Phe Gln Val Cys Leu Cys Gln Asp Glu Val Thr Asp Asp Tyr
 20 25 30

25 Ile Gly Asp Asn Thr Thr Val Asp Tyr Thr Leu Phe Glu Ser Leu Cys
 35 40 45

Ser Lys Lys Asp Val Arg Asn Phe Lys Ala Trp Phe Leu Pro Ile Met
 50 55 60

30 Tyr Ser Ile Ile Cys Phe Val Gly Leu Leu Gly Asn Gly Leu Val Val
 65 70 75 80

Leu Thr Tyr Ile Tyr Phe Lys Arg Leu Lys Thr Met Thr Asp Thr Tyr
 85 90 95

Leu Leu Asn Leu Ala Val Ala Asp Ile Leu Phe Leu Leu Thr Leu Pro

160

	100	105	110
	Phe Trp Ala Tyr Ser Ala Ala Lys Ser Trp Val Phe Gly Val His Phe		
	115	120	125
5	Cys Lys Leu Ile Phe Ala Ile Tyr Lys Met Ser Phe Phe Ser Gly Met		
	130	135	140
	Leu Leu Leu Leu Cys Ile Ser Ile Asp Arg Tyr Val Ala Ile Val Gln		
	145	150	155
	Ala Val Ser Ala His Arg His Arg Ala Arg Val Leu Leu Ile Ser Lys		
	165	170	175
10	Leu Ser Cys Val Gly Ile Trp Ile Leu Ala Thr Val Leu Ser Ile Pro		
	180	185	190
	Glu Leu Leu Tyr Ser Asp Leu Gln Arg Ser Ser Ser Glu Gln Ala Met		
	195	200	205
15	Arg Cys Ser Leu Ile Thr Glu His Val Glu Ala Phe Ile Thr Ile Gln		
	210	215	220
	Val Ala Gln Met Val Ile Gly Phe Leu Val Pro Leu Leu Ala Met Ser		
	225	230	235
	Phe Cys Tyr Leu Val Ile Ile Arg Thr Leu Leu Gln Ala Arg Asn Phe		
	245	250	255
20	Glu Arg Asn Lys Ala Lys Lys Val Ile Ile Ala Val Val Val Val Phe		
	260	265	270
	Ile Val Phe Gln Leu Pro Tyr Asn Gly Val Val Leu Ala Gln Thr Val		
	275	280	285
25	Ala Asn Phe Asn Ile Thr Ser Ser Thr Cys Glu Leu Ser Lys Gln Leu		
	290	295	300
	Asn Ile Ala Tyr Asp Val Thr Tyr Ser Leu Ala Cys Val Arg Cys Cys		
	305	310	315
	Val Asn Pro Phe Leu Tyr Ala Phe Ile Gly Val Lys Phe Arg Asn Asp		
	325	330	335
30	Leu Phe Lys Leu Phe Lys Asp Leu Gly Cys Leu Ser Gln Glu Gln Leu		
	340	345	350
	Arg Gln Trp Ser Ser Cys Arg His Ile Arg Arg Ser Ser Met Ser Val		
	355	360	365
35	Glu Ala Glu Thr Thr Thr Thr Phe Ser Pro		
	370	375	

(206) INFORMATION FOR SEQ ID NO:205:

161

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1086 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:205:

ATGGATATAC AAATGGCAAA CAATTTTACT CCGCCCTCTG CAACTCCTCA GGGAAATGAC 60
TGTGACCTCT ATGCACATCA CAGCACGGCC AGGATAGTAA TGCCTCTGCA TTACAGCCTC 120
10 GTCTTCATCA TTGGGCTCGT GGGAAACTTA CTAGCCTTGG TCGTCATTGT TCAAAACAGG 180
AAAAAATCA ACTCTACCAC CCTCTATTCA ACAAATTTGG TGATTTCTGA TATACTTTTT 240
ACCACGGCTT TGCCTACACG AATAGCCTAC TATGCAATGG GCTTTGACTG GAGAATCGGA 300
GATGCCTTGT GTAGGATAAC TGCGCTAGTG TTTTACATCA ACACATATGC AGGTGTGAAC 360
TTTATGACCT GCCTGAGTAT TGACCGCTTC ATTGCTGTGG TGCACCCTCT ACGCTACAAC 420
15 AAGATAAAAA GGATTGAACA TGCAAAGGC GTGTGCATAT TTGTCTGGAT TCTAGTATTT 480
GCTCAGACAC TCCCACTCCT CATCAACCCT ATGTCAAAGC AGGAGGCTGA AAGGATTACA 540
TGCATGGAGT ATCCAAACTT TGAAGAACT AAATCTCTTC CCTGGATTCT GCTTGGGGCA 600
TGTTTCATAG GATATGTACT TCCACTTATA ATCATTCTCA TCTGCTATTC TCAGATCTGC 660
TGCAAACCTCT TCAGAACTGC CAAACAAAAC CCACTCACTG AGAAATCTGG TGTAACAAA 720
20 AAGGCTAAAA ACACAATTAT TCTTATTATT GTTGTGTTTG TTCTCTGTTT CACACCTTAC 780
CATGTTGCAA TTATTCAACA TATGATTAAG AAGCTTCGTT TCTCTAATTT CCTGGAATGT 840
AGCCAAAGAC ATTCGTTCCA GATTTCTCTG CACTTTACAG TATGCCTGAT GAACTTCAAT 900
TGCTGCATGG ACCCTTTTAT CTACTTCTTT GCATGTAAAG GGTATAAGAG AAAGGTTATG 960
AGGATGCTGA AACGGCAAGT CAGTGTATCG ATTTCTAGTG CTGTGAAGTC AGCCCCTGAA 1020
25 GAAAATTACAC GTGAAATGAC AGAAACGCAG ATGATGATAC ATTCCAAGTC TTCAAATGGA 1080
AAGTGA 1086

(207) INFORMATION FOR SEQ ID NO:206:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 361 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:206:

```

Met Asp Ile Gln Met Ala Asn Asn Phe Thr Pro Pro Ser Ala Thr Pro
1           5           10           15
5  Gln Gly Asn Asp Cys Asp Leu Tyr Ala His His Ser Thr Ala Arg Ile
           20           25           30
Val Met Pro Leu His Tyr Ser Leu Val Phe Ile Ile Gly Leu Val Gly
           35           40           45
10 Asn Leu Leu Ala Leu Val Val Ile Val Gln Asn Arg Lys Lys Ile Asn
           50           55           60
Ser Thr Thr Leu Tyr Ser Thr Asn Leu Val Ile Ser Asp Ile Leu Phe
65           70           75           80
Thr Thr Ala Leu Pro Thr Arg Ile Ala Tyr Tyr Ala Met Gly Phe Asp
           85           90           95
15 Trp Arg Ile Gly Asp Ala Leu Cys Arg Ile Thr Ala Leu Val Phe Tyr
           100          105          110
Ile Asn Thr Tyr Ala Gly Val Asn Phe Met Thr Cys Leu Ser Ile Asp
           115          120          125
20 Arg Phe Ile Ala Val Val His Pro Leu Arg Tyr Asn Lys Ile Lys Arg
           130          135          140
Ile Glu His Ala Lys Gly Val Cys Ile Phe Val Trp Ile Leu Val Phe
145          150          155          160
Ala Gln Thr Leu Pro Leu Leu Ile Asn Pro Met Ser Lys Gln Glu Ala
           165          170          175
25 Glu Arg Ile Thr Cys Met Glu Tyr Pro Asn Phe Glu Glu Thr Lys Ser
           180          185          190
Leu Pro Trp Ile Leu Leu Gly Ala Cys Phe Ile Gly Tyr Val Leu Pro
           195          200          205
30 Leu Ile Ile Ile Leu Ile Cys Tyr Ser Gln Ile Cys Cys Lys Leu Phe
           210          215          220
Arg Thr Ala Lys Gln Asn Pro Leu Thr Glu Lys Ser Gly Val Asn Lys
225          230          235          240
Lys Ala Lys Asn Thr Ile Ile Leu Ile Ile Val Val Phe Val Leu Cys
           245          250          255
35 Phe Thr Pro Tyr His Val Ala Ile Ile Gln His Met Ile Lys Lys Leu
           260          265          270

```

163

Arg Phe Ser Asn Phe Leu Glu Cys Ser Gln Arg His Ser Phe Gln Ile
 275 280 285
 Ser Leu His Phe Thr Val Cys Leu Met Asn Phe Asn Cys Cys Met Asp
 290 295 300
 5 Pro Phe Ile Tyr Phe Phe Ala Cys Lys Gly Tyr Lys Arg Lys Val Met
 305 310 315 320
 Arg Met Leu Lys Arg Gln Val Ser Val Ser Ile Ser Ser Ala Val Lys
 325 330 335
 10 Ser Ala Pro Glu Glu Asn Ser Arg Glu Met Thr Glu Thr Gln Met Met
 340 345 350
 Ile His Ser Lys Ser Ser Asn Gly Lys
 355 360

(208) INFORMATION FOR SEQ ID NO:207:

15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1446 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:207:

ATGCGGTGGC TGTGGCCCCT GGCTGTCTCT CTTGCTGTGA TTTTGGCTGT GGGGCTAAGC 60
 AGGGTCTCTG GGGGTGCCCC CCTGCACCTG GGCAGGCACA GAGCCGAGAC CCAGGAGCAG 120
 CAGAGCCGAT CCAAGAGGGG CACCGAGGAT GAGGAGGCCA AGGGCGTGCA GCAGTATGTG 180
 CCTGAGGAGT GGGCGGAGTA CCCCCGGCCC ATTCACCCTG CTGGCCTGCA GCCAACCAAG 240
 25 CCCTTGGTGG CCACCAGCCC TAACCCCGAC AAGGATGGGG GCACCCCAAG CAGTGGGCAG 300
 GAACTGAGGG GCAATCTGAC AGGGGCACCA GGGCAGAGGC TACAGATCCA GAACCCCTG 360
 TATCCGGTGA CCGAGAGCTC CTACAGTGCC TATGCCATCA TGCTTCTGGC GCTGGTGGTG 420
 TTTGCGGTGG GCATTGTGGG CAACCTGTCTG GTCATGTGCA TCGTGTGGCA CAGCTACTAC 480
 CTGAAGAGCG CCTGGAATC CATCCTTGCC AGCCTGGCCC TCTGGGATTT TCTGGTCCTC 540
 30 TTTTTCTGCC TCCCTATTGT CATCTTCAAC GAGATCACCA AGCAGAGGCT ACTGGGTGAC 600
 GTTCTCTGTC GTGCCGTGCC CTTTCATGGAG GTCTCCTCTC TGGGAGTCAC GACTTTCAGC 660
 CTCTGTGCCC TGGGCATTGA CCGCTTCCAC GTGGCCACCA GCACCCTGCC CAAGGTGAGG 720
 CCCATCGAGC GGTGCCAATC CATCCTGGCC AAGTTGGCTG TCATCTGGGT GGGCTCCATG 780

ACGCTGGCTG TGCCTGAGCT CCTGCTGTGG CAGCTGGCAC AGGAGCCTGC CCCCACCATG 840
 GGCACCCTGG ACTCATGCAT CATGAAACCC TCAGCCAGCC TGCCCGAGTC CCTGTATTCA 900
 CTGGTGATGA CCTACCAGAA CGCCCGCATG TGGTGGTACT TTGGCTGCTA CTTCTGCCTG 960
 CCCATCCTCT TCACAGTCAC CTGCCAGCTG GTGACATGGC GGGTGCGAGG CCCTCCAGGG 1020
 5 AGGAAGTCAG AGTGCAGGGC CAGCAAGCAC GAGCAGTGTG AGAGCCAGCT CAAGAGCACC 1080
 GTGGTGGGCC TGACCGTGGT CTACGCCTTC TGCACCTCC CAGAGAACGT CTGCAACATC 1140
 GTGGTGGCCT ACCTCTCCAC CGAGCTGACC CGCCAGACCC TGGACCTCCT GGGCCTCATC 1200
 AACCAGTTCT CCACCTTCTT CAAGGGCGCC ATCACCCAG TGCTGCTCCT TTGCATCTGC 1260
 AGGCCGCTGG GCCAGGCCTT CCTGGACTGC TGCTGCTGCT GCTGCTGTGA GGAGTGCGGC 1320
 10 GGGGCTTCGG AGGCCTCTGC TGCCAATGGG TCGGACAACA AGCTCAAGAC CGAGGTGTCC 1380
 TCTTCCATCT ACTTCCACAA GCCCAGGGAG TCACCCCCAC TCCTGCCCTT GGGCACACCT 1440
 TGCTGA 1446

(209) INFORMATION FOR SEQ ID NO:208:

- 15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 481 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

- 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:208:

Met Arg Trp Leu Trp Pro Leu Ala Val Ser Leu Ala Val Ile Leu Ala
 1 5 10 15

Val Gly Leu Ser Arg Val Ser Gly Gly Ala Pro Leu His Leu Gly Arg
 20 25 30

- 25 His Arg Ala Glu Thr Gln Glu Gln Gln Ser Arg Ser Lys Arg Gly Thr
 35 40 45

Glu Asp Glu Glu Ala Lys Gly Val Gln Gln Tyr Val Pro Glu Glu Trp
 50 55 60

- 30 Ala Glu Tyr Pro Arg Pro Ile His Pro Ala Gly Leu Gln Pro Thr Lys
 65 70 75 80

Pro Leu Val Ala Thr Ser Pro Asn Pro Asp Lys Asp Gly Gly Thr Pro
 85 90 95

Asp Ser Gly Gln Glu Leu Arg Gly Asn Leu Thr Gly Ala Pro Gly Gln

165

				100					105						110		
	Arg	Leu	Gln	Ile	Gln	Asn	Pro	Leu	Tyr	Pro	Val	Thr	Glu	Ser	Ser	Tyr	
			115						120				125				
5	Ser	Ala	Tyr	Ala	Ile	Met	Leu	Leu	Ala	Leu	Val	Val	Phe	Ala	Val	Gly	
		130					135					140					
	Ile	Val	Gly	Asn	Leu	Ser	Val	Met	Cys	Ile	Val	Trp	His	Ser	Tyr	Tyr	
	145					150					155					160	
	Leu	Lys	Ser	Ala	Trp	Asn	Ser	Ile	Leu	Ala	Ser	Leu	Ala	Leu	Trp	Asp	
					165					170					175		
10	Phe	Leu	Val	Leu	Phe	Phe	Cys	Leu	Pro	Ile	Val	Ile	Phe	Asn	Glu	Ile	
				180					185					190			
	Thr	Lys	Gln	Arg	Leu	Leu	Gly	Asp	Val	Ser	Cys	Arg	Ala	Val	Pro	Phe	
			195					200					205				
15	Met	Glu	Val	Ser	Ser	Leu	Gly	Val	Thr	Thr	Phe	Ser	Leu	Cys	Ala	Leu	
		210					215					220					
	Gly	Ile	Asp	Arg	Phe	His	Val	Ala	Thr	Ser	Thr	Leu	Pro	Lys	Val	Arg	
	225					230					235					240	
	Pro	Ile	Glu	Arg	Cys	Gln	Ser	Ile	Leu	Ala	Lys	Leu	Ala	Val	Ile	Trp	
					245					250					255		
20	Val	Gly	Ser	Met	Thr	Leu	Ala	Val	Pro	Glu	Leu	Leu	Leu	Trp	Gln	Leu	
				260					265					270			
	Ala	Gln	Glu	Pro	Ala	Pro	Thr	Met	Gly	Thr	Leu	Asp	Ser	Cys	Ile	Met	
			275					280					285				
25	Lys	Pro	Ser	Ala	Ser	Leu	Pro	Glu	Ser	Leu	Tyr	Ser	Leu	Val	Met	Thr	
		290					295					300					
	Tyr	Gln	Asn	Ala	Arg	Met	Trp	Trp	Tyr	Phe	Gly	Cys	Tyr	Phe	Cys	Leu	
	305					310					315					320	
	Pro	Ile	Leu	Phe	Thr	Val	Thr	Cys	Gln	Leu	Val	Thr	Trp	Arg	Val	Arg	
					325					330					335		
30	Gly	Pro	Pro	Gly	Arg	Lys	Ser	Glu	Cys	Arg	Ala	Ser	Lys	His	Glu	Gln	
				340					345					350			
	Cys	Glu	Ser	Gln	Leu	Lys	Ser	Thr	Val	Val	Gly	Leu	Thr	Val	Val	Tyr	
			355					360					365				
35	Ala	Phe	Cys	Thr	Leu	Pro	Glu	Asn	Val	Cys	Asn	Ile	Val	Val	Ala	Tyr	
		370					375					380					
	Leu	Ser	Thr	Glu	Leu	Thr	Arg	Gln	Thr	Leu	Asp	Leu	Leu	Gly	Leu	Ile	
	385					390					395					400	

10

15

(ii) MOLECULE TYPE: DNA (genomic)

25

30

167

ACCAAGAAAA TGCTGGCTGT AGTGGTGTTC GCCTTCATCC TCTGCTGGCT CCCCTTCCAC 840
 GTAGGGCGAT ATTTATTTTC CAAATCCTTT GAGCCTGGCT CCTTGGAGAT TGCTCAGATC 900
 AGCCAGTACT GCAACCTCGT GTCCTTTGTC CTCTTCTACC TCAGTGCTGC CATCAACCCC 960
 ATTCTGTACA ACATCATGTC CAAGAAGTAC CGGGTGGCAG TGTTCACTACT TCTGGGATTC 1020
 5 GAACCCTTCT CCCAGAGAAA GCTCTCCACT CTGAAAGATG AAAGTTCTCG GGCCTGGACA 1080
 GAATCTAGTA TTAATACATG A 1101

(211) INFORMATION FOR SEQ ID NO:210:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 366 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:210:

15 Met Trp Asn Ala Thr Pro Ser Glu Glu Pro Gly Phe Asn Leu Thr Leu
 1 5 10 15
 Ala Asp Leu Asp Trp Asp Ala Ser Pro Gly Asn Asp Ser Leu Gly Asp
 20 25 30
 Glu Leu Leu Gln Leu Phe Pro Ala Pro Leu Leu Ala Gly Val Thr Ala
 35 40 45
 Thr Cys Val Ala Leu Phe Val Val Gly Ile Ala Gly Asn Leu Leu Thr
 50 55 60
 Met Leu Val Val Ser Arg Phe Arg Glu Leu Arg Thr Thr Thr Asn Leu
 65 70 75 80
 25 Tyr Leu Ser Ser Met Ala Phe Ser Asp Leu Leu Ile Phe Leu Cys Met
 85 90 95
 Pro Leu Asp Leu Val Arg Leu Trp Gln Tyr Arg Pro Trp Asn Phe Gly
 100 105 110
 Asp Leu Leu Cys Lys Leu Phe Gln Phe Val Ser Glu Ser Cys Thr Tyr
 115 120 125
 Ala Thr Val Leu Thr Ile Thr Ala Leu Ser Val Glu Arg Tyr Phe Ala
 130 135 140
 Ile Cys Phe Pro Leu Arg Ala Lys Val Val Val Thr Lys Gly Arg Val
 145 150 155 160
 35 Lys Leu Val Ile Phe Val Ile Trp Ala Val Ala Phe Cys Ser Ala Gly

		165		170		175										
	Pro	Ile	Phe	Val	Leu	Val	Gly	Val	Glu	His	Glu	Asn	Gly	Thr	Asp	Pro
				180					185					190		
5	Trp	Asp	Thr	Asn	Glu	Cys	Arg	Pro	Thr	Glu	Phe	Ala	Val	Arg	Ser	Gly
			195					200					205			
	Leu	Leu	Thr	Val	Met	Val	Trp	Val	Ser	Ser	Ile	Phe	Phe	Phe	Leu	Pro
		210					215					220				
	Val	Phe	Cys	Leu	Thr	Val	Leu	Tyr	Ser	Leu	Ile	Gly	Arg	Lys	Leu	Trp
	225				230					235					240	
10	Arg	Arg	Arg	Arg	Gly	Asp	Ala	Val	Val	Gly	Ala	Ser	Leu	Arg	Asp	Gln
					245					250				255		
	Asn	His	Lys	Gln	Thr	Lys	Lys	Met	Leu	Ala	Val	Val	Val	Phe	Ala	Phe
			260					265						270		
15	Ile	Leu	Cys	Trp	Leu	Pro	Phe	His	Val	Gly	Arg	Tyr	Leu	Phe	Ser	Lys
		275					280					285				
	Ser	Phe	Glu	Pro	Gly	Ser	Leu	Glu	Ile	Ala	Gln	Ile	Ser	Gln	Tyr	Cys
		290				295				300						
	Asn	Leu	Val	Ser	Phe	Val	Leu	Phe	Tyr	Leu	Ser	Ala	Ala	Ile	Asn	Pro
	305				310					315					320	
20	Ile	Leu	Tyr	Asn	Ile	Met	Ser	Lys	Lys	Tyr	Arg	Val	Ala	Val	Phe	Arg
				325					330					335		
	Leu	Leu	Gly	Phe	Glu	Pro	Phe	Ser	Gln	Arg	Lys	Leu	Ser	Thr	Leu	Lys
			340					345					350			
25	Asp	Glu	Ser	Ser	Arg	Ala	Trp	Thr	Glu	Ser	Ser	Ile	Asn	Thr		
		355					360					365				

(212) INFORMATION FOR SEQ ID NO:211:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1842 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:211:

	ATGCGAGCCC	CGGGCGCGCT	TCTCGCCCGC	ATGTCGCGGC	TACTGCTTCT	GCTACTGCTC	60
35	AAGGTGTCTG	CCTCTTCTGC	CCTCGGGGTC	GCCCCTGCGT	CCAGAAACGA	AACTTGTCTG	120
	GGGGAGAGCT	GTGCACCTAC	AGTGATCCAG	CGCCGCGGCA	GGGACGCCTG	GGGACCGGGA	180

	AATTCCTGCAA	GAGACGTTCT	GCGAGCCCCGA	GCACCCAGGG	AGGAGCAGGG	GGCAGCGTTT	240
	CTTGCGGGAC	CCTCCTGGGA	CCTGCCGGCG	GCCCCGGGGC	GTGACCCGGC	TGCAGGCAGA	300
	GGGGCGGAGG	CGTCGGCAGC	CGGACCCCCG	GGACCTCCAA	CCAGGCCACC	TGGCCCCTGG	360
	AGGTGGAAG	GTGCTCGGGG	TCAGGAGCCT	TCTGAACTT	TGGGGAGAGG	GAACCCACG	420
5	GCCCTCCAGC	TCTTCCTTCA	GATCTCAGAG	GAGGAAGAGA	AGGGTCCCAG	AGGCGCTGGC	480
	ATTTCCGGGC	GTAGCCAGGA	GCAGAGTGTG	AAGACAGTCC	CCGGAGCCAG	CGATCTTTTT	540
	TACTGGCCAA	GGAGAGCCGG	GAAACTCCAG	GGTTCCCACC	ACAAGCCCCT	GTCCAAGACG	600
	GCCAATGGAC	TGGCGGGGCA	CGAAGGGTGG	ACAATTGCAC	TCCCGGGCCG	GGCGCTGGCC	660
	CAGAATGGAT	CCTTGGGTGA	AGGAATCCAT	GAGCCTGGGG	GTCCCCGCCG	GGGAAACAGC	720
10	ACGAACCGGC	GTGTGAGACT	GAAGAACCCC	TTCTACCCGC	TGACCCAGGA	GTCCTATGGA	780
	GCCTACGCGG	TCATGTGTCT	GTCCGTGGTG	ATCTTCGGGA	CCGGCATCAT	TGGCAACCTG	840
	GCGGTGATGT	GCATCGTGTG	CCACAACTAC	TACATGCGGA	GCATCTCCAA	CTCCCTCTTG	900
	GCCAACCTGG	CCTTCTGGGA	CTTTCTCATC	ATCTTCTTCT	GCCTTCCGCT	GGTCATCTTC	960
	CACGAGCTGA	CCAAGAAGTG	GCTGCTGGAG	GACTTCTCCT	GCAAGATCGT	GCCCTATATA	1020
15	GAGGTGCGCT	CTCTGGGAGT	CACCACTTTC	ACCTTATGTG	CTCTGTGCAT	AGACCGCTTC	1080
	CGTGCTGCCA	CCAACGTACA	GATGTACTAC	GAAATGATCG	AAAATTGTTC	CTCAACAAC	1140
	GCCAAACTTG	CTGTTATATG	GGTGGGAGCT	CTATTGTTAG	CACTTCCAGA	AGTTGTTCTC	1200
	CGCCAGCTGA	GCAAGGAGGA	TTTGGGGTTT	AGTGGCCGAG	CTCCGGCAGA	AAGGTGCATT	1260
	ATTAAGATCT	CTCCTGATTT	ACCAGACACC	ATCTATGTTC	TAGCCCTCAC	CTACGACAGT	1320
20	GCGAGACTGT	GGTGGTATTT	TGGCTGTTAC	TTTTGTTTGC	CCACGCTTTT	CACCATCACC	1380
	TGCTCTCTAG	TGACTGCGAG	GAAAATCCGC	AAAGCAGAGA	AAGCCTGTAC	CCGAGGGAAT	1440
	AAACGGCAGA	TTCAACTAGA	GAGTCAGATG	AAGTGACAG	TAGTGGCACT	GACCATTTTA	1500
	TATGGATTTT	GCATTATTCC	TGAAAATATC	TGCAACATTG	TTACTGCCTA	CATGGCTACA	1560
	GGGGTTTCAC	AGCAGACAAT	GGACCTCCTT	AATATCATCA	GCCAGTTCCT	TTTGTTCTTT	1620
25	AAGTCCTGTG	TCACCCAGT	CCTCCTTTTC	TGTCTCTGCA	AACCCTTCAG	TCGGGCCTTC	1680
	ATGGAGTGCT	GCTGCTGTTG	CTGTGAGGAA	TGCATTGAGA	AGTCTTCAAC	GGTGACCAGT	1740
	GATGACAATG	ACAACGAGTA	CACCACGGAA	CTCGAACTCT	CGCCTTTCAG	TACCATACGC	1800
	CGTGAAATGT	CCACTTTTGC	TTCTGTGCGA	ACTCATTGCT	GA		1842

(213) INFORMATION FOR SEQ ID NO:212:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 613 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:212:

10 Met Arg Ala Pro Gly Ala Leu Leu Ala Arg Met Ser Arg Leu Leu Leu
 1 5 10 15
 Leu Leu Leu Leu Lys Val Ser Ala Ser Ser Ala Leu Gly Val Ala Pro
 20 25 30
 Ala Ser Arg Asn Glu Thr Cys Leu Gly Glu Ser Cys Ala Pro Thr Val
 35 40 45
 15 Ile Gln Arg Arg Gly Arg Asp Ala Trp Gly Pro Gly Asn Ser Ala Arg
 50 55 60
 Asp Val Leu Arg Ala Arg Ala Pro Arg Glu Glu Gln Gly Ala Ala Phe
 65 70 75 80
 20 Leu Ala Gly Pro Ser Trp Asp Leu Pro Ala Ala Pro Gly Arg Asp Pro
 85 90 95
 Ala Ala Gly Arg Gly Ala Glu Ala Ser Ala Ala Gly Pro Pro Gly Pro
 100 105 110
 Pro Thr Arg Pro Pro Gly Pro Trp Arg Trp Lys Gly Ala Arg Gly Gln
 115 120 125
 25 Glu Pro Ser Glu Thr Leu Gly Arg Gly Asn Pro Thr Ala Leu Gln Leu
 130 135 140
 Phe Leu Gln Ile Ser Glu Glu Glu Glu Lys Gly Pro Arg Gly Ala Gly
 145 150 155 160
 30 Ile Ser Gly Arg Ser Gln Glu Gln Ser Val Lys Thr Val Pro Gly Ala
 165 170 175
 Ser Asp Leu Phe Tyr Trp Pro Arg Arg Ala Gly Lys Leu Gln Gly Ser
 180 185 190
 His His Lys Pro Leu Ser Lys Thr Ala Asn Gly Leu Ala Gly His Glu
 195 200 205
 35 Gly Trp Thr Ile Ala Leu Pro Gly Arg Ala Leu Ala Gln Asn Gly Ser
 210 215 220
 Leu Gly Glu Gly Ile His Glu Pro Gly Gly Pro Arg Arg Gly Asn Ser

171

	225		230		235		240
	Thr Asn Arg Arg Val Arg Leu Lys Asn Pro Phe Tyr Pro Leu Thr Gln						
		245			250		255
5	Glu Ser Tyr Gly Ala Tyr Ala Val Met Cys Leu Ser Val Val Ile Phe						
		260		265		270	
	Gly Thr Gly Ile Ile Gly Asn Leu Ala Val Met Cys Ile Val Cys His						
		275		280		285	
	Asn Tyr Tyr Met Arg Ser Ile Ser Asn Ser Leu Leu Ala Asn Leu Ala						
		290		295		300	
10	Phe Trp Asp Phe Leu Ile Ile Phe Phe Cys Leu Pro Leu Val Ile Phe						
		305		310		315	320
	His Glu Leu Thr Lys Lys Trp Leu Leu Glu Asp Phe Ser Cys Lys Ile						
			325		330		335
15	Val Pro Tyr Ile Glu Val Ala Ser Leu Gly Val Thr Thr Phe Thr Leu						
		340		345			350
	Cys Ala Leu Cys Ile Asp Arg Phe Arg Ala Ala Thr Asn Val Gln Met						
		355		360		365	
	Tyr Tyr Glu Met Ile Glu Asn Cys Ser Ser Thr Thr Ala Lys Leu Ala						
		370		375		380	
20	Val Ile Trp Val Gly Ala Leu Leu Leu Ala Leu Pro Glu Val Val Leu						
		385		390		395	400
	Arg Gln Leu Ser Lys Glu Asp Leu Gly Phe Ser Gly Arg Ala Pro Ala						
			405		410		415
25	Glu Arg Cys Ile Ile Lys Ile Ser Pro Asp Leu Pro Asp Thr Ile Tyr						
		420		425		430	
	Val Leu Ala Leu Thr Tyr Asp Ser Ala Arg Leu Trp Trp Tyr Phe Gly						
		435		440		445	
	Cys Tyr Phe Cys Leu Pro Thr Leu Phe Thr Ile Thr Cys Ser Leu Val						
		450		455		460	
30	Thr Ala Arg Lys Ile Arg Lys Ala Glu Lys Ala Cys Thr Arg Gly Asn						
		465		470		475	480
	Lys Arg Gln Ile Gln Leu Glu Ser Gln Met Lys Cys Thr Val Val Ala						
			485		490		495
35	Leu Thr Ile Leu Tyr Gly Phe Cys Ile Ile Pro Glu Asn Ile Cys Asn						
		500		505		510	
	Ile Val Thr Ala Tyr Met Ala Thr Gly Val Ser Gln Gln Thr Met Asp						
		515		520		525	

Leu Leu Asn Ile Ile Ser Gln Phe Leu Leu Phe Phe Lys Ser Cys Val
 530 535 540
 Thr Pro Val Leu Leu Phe Cys Leu Cys Lys Pro Phe Ser Arg Ala Phe
 545 550 555 560
 5 Met Glu Cys Cys Cys Cys Cys Glu Glu Cys Ile Gln Lys Ser Ser
 565 570 575
 Thr Val Thr Ser Asp Asp Asn Asp Asn Glu Tyr Thr Thr Glu Leu Glu
 580 585 590
 10 Leu Ser Pro Phe Ser Thr Ile Arg Arg Glu Met Ser Thr Phe Ala Ser
 595 600 605
 Val Gly Thr His Cys
 610

(214) INFORMATION FOR SEQ ID NO:213:

15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1248 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:213:

ATGGTTTTTG CTCACAGAAT GGATAACAGC AAGCCACATT TGATTATTCC TAACTTCTG 60
 GTGCCCCTCC AAAACCGCAG CTGCACTGAA ACAGCCACAC CTCTGCCAAG CCAATACCTG 120
 ATGGAATTAA GTGAGGAGCA CAGTTGGATG AGCAACCAAA CAGACCTTCA CTATGTGCTG 180
 AAACCCGGGG AAGTGGCCAC AGCCAGCATC TTCTTTGGGA TTCTGTGGTT GTTTTCTATC 240
 25 TTCGGCAATT CCCTGGTTTG TTTGGTCATC CATAGGAGTA GGAGGACTCA GTCTACCACC 300
 AACTACTTTG TGGTCTCCAT GGCATGTGCT GACCTTCTCA TCAGCGTTGC CAGCAGCCCT 360
 TTCGTCCTGC TCCAGTTCAC CACTGGAAGG TGGACGCTGG GTAGTGCAAC GTGCAAGGTT 420
 GTGCGATATT TTCAATATCT CACTCCAGGT GTCCAGATCT ACGTTCTCCT CTCCATCTGC 480
 ATAGACCGGT TCTACACCAT CGTCTATCCT CTGAGCTTCA AGGTGTCCAG AGAAAAAGCC 540
 30 AAGAAAATGA TTGCGGCATC GTGGATCTTT GATGCAGGCT TTGTGACCCC TGTGCTCTTT 600
 TTCTATGGCT CCAACTGGGA CAGTCATTGT AACTATTTC TCCCCTCCTC TTGGGAAGGC 660
 ACTGCCTACA CTGTCATCCA CTTCTTGGTG GGCTTTGTGA TTCCATCTGT CCTCATAATT 720
 TTATTTTACC AAAAGGTCAT AAAATATATT TGGAGAATAG GCACAGATGG CCGAACGGTG 780

173

AGGAGGACAA TGAACATTGT CCCTCGGACA AAAGTGAAAA CTAAAAAGAT GTTCCTCATT 840
 TTAAATCTGT TGTTTTTGCT CTCCTGGCTG CCTTTTCATG TAGCTCAGCT ATGGCACCCC 900
 CATGAACAAG ACTATAAGAA AAGTTCCCTT GTTTTCACAG CTATCACATG GATATCCTTT 960
 AGTTCTTCAG CCTCTAAACC TACTCTGTAT TCAATTTATA ATGCCAATTT TCGGAGAGGG 1020
 5 ATGAAAGAGA CTTTTTGCAT GTCCTCTATG AAATGTTACC GAAGCAATGC CTATACTATC 1080
 ACAACAAGTT CAAGGATGGC CAAAAAAAAC TACGTTGGCA TTTCAGAAAT CCCTTCCATG 1140
 GCCAAACTA TTACCAAAGA CTCGATCTAT GACTCATTTG ACAGAGAAGC CAAGGAAAAA 1200
 AAGCTTGCTT GGCCCATTA CTCAAATCCA CCAAATACTT TTGTCTAA 1248

(215) INFORMATION FOR SEQ ID NO:214:

10 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 415 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

15 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:214:

Met Val Phe Ala His Arg Met Asp Asn Ser Lys Pro His Leu Ile Ile
 1 5 10 15
 Pro Thr Leu Leu Val Pro Leu Gln Asn Arg Ser Cys Thr Glu Thr Ala
 20 20 25 30
 Thr Pro Leu Pro Ser Gln Tyr Leu Met Glu Leu Ser Glu Glu His Ser
 35 40 45
 Trp Met Ser Asn Gln Thr Asp Leu His Tyr Val Leu Lys Pro Gly Glu
 50 55 60
 25 Val Ala Thr Ala Ser Ile Phe Phe Gly Ile Leu Trp Leu Phe Ser Ile
 65 70 75 80
 Phe Gly Asn Ser Leu Val Cys Leu Val Ile His Arg Ser Arg Arg Thr
 85 90 95
 30 Gln Ser Thr Thr Asn Tyr Phe Val Val Ser Met Ala Cys Ala Asp Leu
 100 105 110
 Leu Ile Ser Val Ala Ser Thr Pro Phe Val Leu Leu Gln Phe Thr Thr
 115 120 125
 Gly Arg Trp Thr Leu Gly Ser Ala Thr Cys Lys Val Val Arg Tyr Phe
 130 135 140

Gln Tyr Leu Thr Pro Gly Val Gln Ile Tyr Val Leu Leu Ser Ile Cys
 145 150 155 160
 Ile Asp Arg Phe Tyr Thr Ile Val Tyr Pro Leu Ser Phe Lys Val Ser
 165 170 175
 5 Arg Glu Lys Ala Lys Lys Met Ile Ala Ala Ser Trp Ile Phe Asp Ala
 180 185 190
 Gly Phe Val Thr Pro Val Leu Phe Phe Tyr Gly Ser Asn Trp Asp Ser
 195 200 205
 10 His Cys Asn Tyr Phe Leu Pro Ser Ser Trp Glu Gly Thr Ala Tyr Thr
 210 215 220
 Val Ile His Phe Leu Val Gly Phe Val Ile Pro Ser Val Leu Ile Ile
 225 230 235 240
 Leu Phe Tyr Gln Lys Val Ile Lys Tyr Ile Trp Arg Ile Gly Thr Asp
 245 250 255
 15 Gly Arg Thr Val Arg Arg Thr Met Asn Ile Val Pro Arg Thr Lys Val
 260 265 270
 Lys Thr Lys Lys Met Phe Leu Ile Leu Asn Leu Leu Phe Leu Leu Ser
 275 280 285
 20 Trp Leu Pro Phe His Val Ala Gln Leu Trp His Pro His Glu Gln Asp
 290 295 300
 Tyr Lys Lys Ser Ser Leu Val Phe Thr Ala Ile Thr Trp Ile Ser Phe
 305 310 315 320
 Ser Ser Ser Ala Ser Lys Pro Thr Leu Tyr Ser Ile Tyr Asn Ala Asn
 325 330 335
 25 Phe Arg Arg Gly Met Lys Glu Thr Phe Cys Met Ser Ser Met Lys Cys
 340 345 350
 Tyr Arg Ser Asn Ala Tyr Thr Ile Thr Thr Ser Ser Arg Met Ala Lys
 355 360 365
 30 Lys Asn Tyr Val Gly Ile Ser Glu Ile Pro Ser Met Ala Lys Thr Ile
 370 375 380
 Thr Lys Asp Ser Ile Tyr Asp Ser Phe Asp Arg Glu Ala Lys Glu Lys
 385 390 395 400
 Lys Leu Ala Trp Pro Ile Asn Ser Asn Pro Pro Asn Thr Phe Val
 405 410 415

35 (216) INFORMATION FOR SEQ ID NO:215:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1842 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:215:

ATGGGGCCCA	CCCTAGCGGT	TCCCACCCCC	TATGGCTGTA	TTGGCTGTAA	GCTACCCCAG	60
CCAGAATACC	CACCGGCTCT	AATCATCTTT	ATGTTCTGCG	CGATGGTTAT	CACCATCGTT	120
GTAGACCTAA	TCGGCAACTC	CATGGTCATT	TTGGCTGTGA	CGAAGAACAA	GAAGCTCCGG	180
AATTCTGGCA	ACATCTTCGT	GGTCAGTCTC	TCTGTGGCCG	ATATGCTGGT	GGCCATCTAC	240
10 CCATACCCTT	TGATGCTGCA	TGCCATGTCC	ATTGGGGGCT	GGGATCTGAG	CCAGTTACAG	300
TGCCAGATGG	TCGGGTTTCA	CACAGGGCTG	AGTGTGGTCG	GCTCCATCTT	CAACATCGTG	360
GCAATCGCTA	TCAACCGTTA	CTGCTACATC	TGCCACAGCC	TCCAGTACGA	ACGGATCTTC	420
AGTGTGCGCA	ATACCTGCAT	CTACCTGGTC	ATCACCTGGA	TCATGACCGT	CCTGGCTGTC	480
CTGCCCAACA	TGTACATTGG	CACCATCGAG	TACGATCCTC	GCACCTACAC	CTGCATCTTC	540
15 AACTATCTGA	ACAACCCTGT	CTTCACTGTT	ACCATCGTCT	GCATCCACTT	CGTCCTCCCT	600
CTCCTCATCG	TGGGTTTCTG	CTACGTGAGG	ATCTGGACCA	AAGTGCTGGC	GGCCCGTGAC	660
CCTGCAGGGC	AGAATCCTGA	CAACCAACTT	GCTGAGGTTT	GCAATAAACT	AACCATGTTT	720
GTGATCTTCC	TCCTCTTTGC	AGTGTGCTGG	TGCCCTATCA	ACGTGCTCAC	TGTCTTGGTG	780
GCTGTCACTC	CGAAGGAGAT	GGCAGGCAAG	ATCCCCAACT	GGCTTTATCT	TGCAGCCTAC	840
20 TTCATAGCCT	ACTTCAACAG	CTGCCTCAAC	GCTGTGATCT	ACGGGCTCCT	CAATGAGAAT	900
TTCCGAAGAG	AATACTGGAC	CATCTTCCAT	GCTATGCGGC	ACCCTATCAT	ATTCTTCTCT	960
GGCCTCATCA	GTGATATTCT	TGAGATGCAG	GAGGCCCGTA	CCCTGGCCCG	CGCCCGTGCC	1020
CATGCTCGCG	ACCAAGCTCG	TGAACAAGAC	CGTGCCCATG	CCTGTCCTGC	TGTGGAGGAA	1080
ACCCCGATGA	ATGTCCGGAA	TGTTCCATTA	CCTGGTGATG	CTGCAGCTGG	CCACCCGAC	1140
25 CGTGCCTCTG	GCCACCCTAA	GCCCCATTCC	AGATCCTCCT	CTGCCTATCG	CAAATCTGCC	1200
TCTACCCACC	ACAAGTCTGT	CTTTAGCCAC	TCCAAGGCTG	CCTCTGGTCA	CCTCAAGCCT	1260
GTCTCTGGCC	ACTCCAAGCC	TGCCTCTGGT	CACCCCAAGT	CTGCCACTGT	CTACCCTAAG	1320
CCTGCCTCTG	TCCATTTCAA	GGCTGACTCT	GTCCATTTC	AGGGTGACTC	TGTCCATTTC	1380
AAGCCTGACT	CTGTTTATTT	CAAGCCTGCT	TCCAGCAACC	CCAAGCCCAT	CACTGGCCAC	1440

CATGTCTCTG CTGGCAGCCA CTCCAAGTCT GCCTTCAATG CTGCCACCAG CCACCCTAAA 1500
 CCCATCAAGC CAGCTACCAG CCATGCTGAG CCCACCACTG CTGACTATCC CAAGCCTGCC 1560
 ACTACCAGCC ACCCTAAGCC CGCTGCTGCT GACAACCCTG AGCTCTCTGC CTCCCATTGC 1620
 CCCGAGATCC CTGCCATTGC CCACCCTGTG TCTGACGACA GTGACCTCCC TGAGTCGGCC 1680
 5 TCTAGCCCTG CCGCTGGGCC CACCAAGCCT GCTGCCAGCC AGCTGGAGTC TGACACCATC 1740
 GCTGACCTTC CTGACCCTAC TGTAGTCACT ACCAGTACCA ATGATTACCA TGATGTCGTG 1800
 GTTGTGATG TTGAAGATGA TCCTGATGAA ATGGCTGTGT GA 1842

(217) INFORMATION FOR SEQ ID NO:216:

10 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 613 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

 (ii) MOLECULE TYPE: protein

 15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:216:

 Met Gly Pro Thr Leu Ala Val Pro Thr Pro Tyr Gly Cys Ile Gly Cys
 1 5 10 15
 Lys Leu Pro Gln Pro Glu Tyr Pro Pro Ala Leu Ile Ile Phe Met Phe
 20 25 30
 20 Cys Ala Met Val Ile Thr Ile Val Val Asp Leu Ile Gly Asn Ser Met
 35 40 45
 Val Ile Leu Ala Val Thr Lys Asn Lys Lys Leu Arg Asn Ser Gly Asn
 50 55 60
 25 Ile Phe Val Val Ser Leu Ser Val Ala Asp Met Leu Val Ala Ile Tyr
 65 70 75 80
 Pro Tyr Pro Leu Met Leu His Ala Met Ser Ile Gly Gly Trp Asp Leu
 85 90 95
 Ser Gln Leu Gln Cys Gln Met Val Gly Phe Ile Thr Gly Leu Ser Val
 100 105 110
 30 Val Gly Ser Ile Phe Asn Ile Val Ala Ile Ala Ile Asn Arg Tyr Cys
 115 120 125
 Tyr Ile Cys His Ser Leu Gln Tyr Glu Arg Ile Phe Ser Val Arg Asn
 130 135 140
 35 Thr Cys Ile Tyr Leu Val Ile Thr Trp Ile Met Thr Val Leu Ala Val
 145 150 155 160

177

	Leu	Pro	Asn	Met	Tyr	Ile	Gly	Thr	Ile	Glu	Tyr	Asp	Pro	Arg	Thr	Tyr	165	170	175
	Thr	Cys	Ile	Phe	Asn	Tyr	Leu	Asn	Asn	Pro	Val	Phe	Thr	Val	Thr	Ile	180	185	190
5	Val	Cys	Ile	His	Phe	Val	Leu	Pro	Leu	Leu	Ile	Val	Gly	Phe	Cys	Tyr	195	200	205
	Val	Arg	Ile	Trp	Thr	Lys	Val	Leu	Ala	Ala	Arg	Asp	Pro	Ala	Gly	Gln	210	215	220
10	Asn	Pro	Asp	Asn	Gln	Leu	Ala	Glu	Val	Arg	Asn	Lys	Leu	Thr	Met	Phe	225	230	235
	Val	Ile	Phe	Leu	Leu	Phe	Ala	Val	Cys	Trp	Cys	Pro	Ile	Asn	Val	Leu	245	250	255
	Thr	Val	Leu	Val	Ala	Val	Ser	Pro	Lys	Glu	Met	Ala	Gly	Lys	Ile	Pro	260	265	270
15	Asn	Trp	Leu	Tyr	Leu	Ala	Ala	Tyr	Phe	Ile	Ala	Tyr	Phe	Asn	Ser	Cys	275	280	285
	Leu	Asn	Ala	Val	Ile	Tyr	Gly	Leu	Leu	Asn	Glu	Asn	Phe	Arg	Arg	Glu	290	295	300
20	Tyr	Trp	Thr	Ile	Phe	His	Ala	Met	Arg	His	Pro	Ile	Ile	Phe	Phe	Ser	305	310	315
	Gly	Leu	Ile	Ser	Asp	Ile	Arg	Glu	Met	Gln	Glu	Ala	Arg	Thr	Leu	Ala	325	330	335
	Arg	Ala	Arg	Ala	His	Ala	Arg	Asp	Gln	Ala	Arg	Glu	Gln	Asp	Arg	Ala	340	345	350
25	His	Ala	Cys	Pro	Ala	Val	Glu	Glu	Thr	Pro	Met	Asn	Val	Arg	Asn	Val	355	360	365
	Pro	Leu	Pro	Gly	Asp	Ala	Ala	Ala	Gly	His	Pro	Asp	Arg	Ala	Ser	Gly	370	375	380
30	His	Pro	Lys	Pro	His	Ser	Arg	Ser	Ser	Ser	Ala	Tyr	Arg	Lys	Ser	Ala	385	390	395
	Ser	Thr	His	His	Lys	Ser	Val	Phe	Ser	His	Ser	Lys	Ala	Ala	Ser	Gly	405	410	415
	His	Leu	Lys	Pro	Val	Ser	Gly	His	Ser	Lys	Pro	Ala	Ser	Gly	His	Pro	420	425	430
35	Lys	Ser	Ala	Thr	Val	Tyr	Pro	Lys	Pro	Ala	Ser	Val	His	Phe	Lys	Ala	435	440	445
	Asp	Ser	Val	His	Phe	Lys	Gly	Asp	Ser	Val	His	Phe	Lys	Pro	Asp	Ser			

	450		455		460
	Val His Phe Lys Pro Ala Ser Ser Asn Pro Lys Pro Ile Thr Gly His				
	465		470		475 480
5	His Val Ser Ala Gly Ser His Ser Lys Ser Ala Phe Asn Ala Ala Thr				
		485		490	495
	Ser His Pro Lys Pro Ile Lys Pro Ala Thr Ser His Ala Glu Pro Thr				
		500		505	510
	Thr Ala Asp Tyr Pro Lys Pro Ala Thr Thr Ser His Pro Lys Pro Ala				
		515		520	525
10	Ala Ala Asp Asn Pro Glu Leu Ser Ala Ser His Cys Pro Glu Ile Pro				
		530		535	540
	Ala Ile Ala His Pro Val Ser Asp Asp Ser Asp Leu Pro Glu Ser Ala				
		545	550	555	560
15	Ser Ser Pro Ala Ala Gly Pro Thr Lys Pro Ala Ala Ser Gln Leu Glu				
		565		570	575
	Ser Asp Thr Ile Ala Asp Leu Pro Asp Pro Thr Val Val Thr Thr Ser				
		580		585	590
	Thr Asn Asp Tyr His Asp Val Val Val Val Asp Val Glu Asp Asp Pro				
		595		600	605
20	Asp Glu Met Ala Val				
		610			

(218) INFORMATION FOR SEQ ID NO:217:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1854 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:217:
- ATGGGGCCCA CCCTAGCGGT TCCCACCCCC TATGGCTGTA TTGGCTGTAA GCTACCCCAG
- CCAGAATACC CACCGGCTCT AATCATCTTT ATGTTCTGCG CGATGGTTAT CACCATCGTT
- GTAGACCTAA TCGGCAACTC CATGGTCATT TTGGCTGTGA CGAAGAACAA GAAGCTCCGG
- AATTCTGGCA ACATCTTCGT GGTCAGTCTC TCTGTGGCCG ATATGCTGGT GGCCATCTAC
- CCATACCCTT TGATGCTGCA TGCCATGTCC ATTGGGGGCT GGGATCTGAG CCAGTTACAG
- TGCCAGATGG TCGGGTTCAT CACAGGGCTG AGTGTGGTCG GCTCCATCTT CAACATCGTG

	GCAATCGCTA	TCAACCGTTA	CTGCTACATC	TGCCACAGCC	TCCAGTACGA	ACGGATCTTC	420
	AGTGTGCGCA	ATACCTGCAT	CTACCTGGTC	ATCACCTGGA	TCATGACCGT	CCTGGCTGTC	480
	CTGCCCCAACA	TGTACATTGG	CACCATCGAG	TACGATCCTC	GCACCTACAC	CTGCATCTTC	540
	AACTATCTGA	ACAACCCTGT	CTTCACTGTT	ACCATCGTCT	GCATCCACTT	CGTCCTCCCT	600
5	CTCCTCATCG	TGGGTTTCTG	CTACGTGAGG	ATCTGGACCA	AAGTGCTGGC	GGCCCGTGAC	660
	CCTGCAGGGC	AGAATCCTGA	CAACCAACTT	GCTGAGGTTC	GCAATAAACT	AACCATGTTT	720
	GTGATCTTCC	TCCTCTTTGC	AGTGTGCTGG	TGCCCTATCA	ACGTGCTCAC	TGTCTTGGTG	780
	GCTGTCAGTC	CGAAGGAGAT	GGCAGGCAAG	ATCGCCAACT	GGCTTTATCT	TGCAGCCTAC	840
	TTCATAGCCT	ACTTCAACAG	CTGCCTCAAC	GCTGTGATCT	ACGGGCTCCT	CAATGAGAAT	900
10	TTCCGAAGAG	AATACTGGAC	CATCTTCCAT	GCTATGCGGC	ACCCTATCAT	ATTCTTCTCT	960
	GGCCTCATCA	GTGATATTCG	TGAGATGCAG	GAGGCCCGTA	CCCTGGCCCCG	CGCCCGTGCC	1020
	CATGCTCGCG	ACCAAGCTCG	TGAACAAGAC	CGTGCCCATG	CCTGTCCTGC	TGTGGAGGAA	1080
	ACCCCGATGA	ATGTCCGGAA	TGTTCCATTA	CCTGGTGATG	CTGCAGCTGG	CCACCCCGAC	1140
	CGTGCTCTG	GCCACCCTAA	GCCCCATTCC	AGATCCTCCT	CTGCCTATCG	CAAATCTGCC	1200
15	TCTACCCACC	ACAAGTCTGT	CTTTAGCCAC	TCCAAGGCTG	CCTCTGGTCA	CCTCAAGCCT	1260
	GTCTCTGGCC	ACTCCAAGCC	TGCCTCTGGT	CACCCCAAGT	CTGCCACTGT	CTACCCTAAG	1320
	CCTGCCTCTG	TCCATTTCAA	GGCTGACTCT	GTCCATTTCA	AGGGTGACTC	TGTCCATTTC	1380
	AAGCCTGACT	CTGTTCAATT	CAAGCCTGCT	TCCAGCAACC	CCAAGCCCAT	CACTGGCCAC	1440
	CATGTCTCTG	CTGGCAGCCA	CTCCAAGTCT	GCCTTCAGTG	CTGCCACCAG	CCACCCTAAA	1500
20	CCCACCACTG	GCCACATCAA	GCCAGCTACC	AGCCATGCTG	AGCCCACCAC	TGCTGACTAT	1560
	CCCAAGCCTG	CCACTACCAG	CCACCCTAAG	CCCCTGCTG	CTGACAACCC	TGAGCTCTCT	1620
	GCCTCCCATT	GCCCCGAGAT	CCCTGCCATT	GCCCACCCTG	TGTCTGACGA	CAGTGACCTC	1680
	CCTGAGTCGG	CCTCTAGCCC	TGCCGCTGGG	CCCACCAAGC	CTGCTGCCAG	CCAGCTGGAG	1740
	TCTGACACCA	TCGCTGACCT	TCCTGACCCT	ACTGTAGTCA	CTACCAGTAC	CAATGATTAC	1800
25	CATGATGTCG	TGGTTGTTGA	TGTTGAAGAT	GATCCTGATG	AAATGGCTGT	GTGA	1854

(219) INFORMATION FOR SEQ ID NO:218:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 617 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:218:

5 Met Gly Pro Thr Leu Ala Val Pro Thr Pro Tyr Gly Cys Ile Gly Cys
 1 5 10 15
 Lys Leu Pro Gln Pro Glu Tyr Pro Pro Ala Leu Ile Ile Phe Met Phe
 20 25 30
 10 Cys Ala Met Val Ile Thr Ile Val Val Asp Leu Ile Gly Asn Ser Met
 35 40 45
 Val Ile Leu Ala Val Thr Lys Asn Lys Lys Leu Arg Asn Ser Gly Asn
 50 55 60
 Ile Phe Val Val Ser Leu Ser Val Ala Asp Met Leu Val Ala Ile Tyr
 65 70 75 80
 15 Pro Tyr Pro Leu Met Leu His Ala Met Ser Ile Gly Gly Trp Asp Leu
 85 90 95
 Ser Gln Leu Gln Cys Gln Met Val Gly Phe Ile Thr Gly Leu Ser Val
 100 105 110
 20 Val Gly Ser Ile Phe Asn Ile Val Ala Ile Ala Ile Asn Arg Tyr Cys
 115 120 125
 Tyr Ile Cys His Ser Leu Gln Tyr Glu Arg Ile Phe Ser Val Arg Asn
 130 135 140
 Thr Cys Ile Tyr Leu Val Ile Thr Trp Ile Met Thr Val Leu Ala Val
 145 150 155 160
 25 Leu Pro Asn Met Tyr Ile Gly Thr Ile Glu Tyr Asp Pro Arg Thr Tyr
 165 170 175
 Thr Cys Ile Phe Asn Tyr Leu Asn Asn Pro Val Phe Thr Val Thr Ile
 180 185 190
 30 Val Cys Ile His Phe Val Leu Pro Leu Leu Ile Val Gly Phe Cys Tyr
 195 200 205
 Val Arg Ile Trp Thr Lys Val Leu Ala Ala Arg Asp Pro Ala Gly Gln
 210 215 220
 Asn Pro Asp Asn Gln Leu Ala Glu Val Arg Asn Lys Leu Thr Met Phe
 225 230 235 240
 35 Val Ile Phe Leu Leu Phe Ala Val Cys Trp Cys Pro Ile Asn Val Leu
 245 250 255

181

	Thr	Val	Leu	Val	Ala	Val	Ser	Pro	Lys	Glu	Met	Ala	Gly	Lys	Ile	Pro	
				260					265					270			
	Asn	Trp	Leu	Tyr	Leu	Ala	Ala	Tyr	Phe	Ile	Ala	Tyr	Phe	Asn	Ser	Cys	
			275					280					285				
5	Leu	Asn	Ala	Val	Ile	Tyr	Gly	Leu	Leu	Asn	Glu	Asn	Phe	Arg	Arg	Glu	
		290					295					300					
	Tyr	Trp	Thr	Ile	Phe	His	Ala	Met	Arg	His	Pro	Ile	Ile	Phe	Phe	Ser	
	305					310					315					320	
10	Gly	Leu	Ile	Ser	Asp	Ile	Arg	Glu	Met	Gln	Glu	Ala	Arg	Thr	Leu	Ala	
					325					330					335		
	Arg	Ala	Arg	Ala	His	Ala	Arg	Asp	Gln	Ala	Arg	Glu	Gln	Asp	Arg	Ala	
				340					345					350			
	His	Ala	Cys	Pro	Ala	Val	Glu	Glu	Thr	Pro	Met	Asn	Val	Arg	Asn	Val	
			355					360					365				
15	Pro	Leu	Pro	Gly	Asp	Ala	Ala	Ala	Gly	His	Pro	Asp	Arg	Ala	Ser	Gly	
		370					375					380					
	His	Pro	Lys	Pro	His	Ser	Arg	Ser	Ser	Ser	Ala	Tyr	Arg	Lys	Ser	Ala	
	385					390					395					400	
20	Ser	Thr	His	His	Lys	Ser	Val	Phe	Ser	His	Ser	Lys	Ala	Ala	Ser	Gly	
					405					410					415		
	His	Leu	Lys	Pro	Val	Ser	Gly	His	Ser	Lys	Pro	Ala	Ser	Gly	His	Pro	
				420					425					430			
	Lys	Ser	Ala	Thr	Val	Tyr	Pro	Lys	Pro	Ala	Ser	Val	His	Phe	Lys	Ala	
			435					440					445				
25	Asp	Ser	Val	His	Phe	Lys	Gly	Asp	Ser	Val	His	Phe	Lys	Pro	Asp	Ser	
		450					455					460					
	Val	His	Phe	Lys	Pro	Ala	Ser	Ser	Asn	Pro	Lys	Pro	Ile	Thr	Gly	His	
	465					470					475					480	
30	His	Val	Ser	Ala	Gly	Ser	His	Ser	Lys	Ser	Ala	Phe	Ser	Ala	Ala	Thr	
					485					490					495		
	Ser	His	Pro	Lys	Pro	Thr	Thr	Gly	His	Ile	Lys	Pro	Ala	Thr	Ser	His	
				500					505					510			
	Ala	Glu	Pro	Thr	Thr	Ala	Asp	Tyr	Pro	Lys	Pro	Ala	Thr	Thr	Ser	His	
				515				520					525				
35	Pro	Lys	Pro	Thr	Ala	Ala	Asp	Asn	Pro	Glu	Leu	Ser	Ala	Ser	His	Cys	
		530					535					540					
	Pro	Glu	Ile	Pro	Ala	Ile	Ala	His	Pro	Val	Ser	Asp	Asp	Ser	Asp	Leu	

RECEIVED - INO MAY 21 1964

TTGCTGCTCA CTGCTGTTTG GCTGCCCAA GTCTCCCTGC TGGCAAACCC TGTTCCTTTT 960
 CTTACTGTGA ACAAATCTGT CCGCAAGTGC TTGATAGGGA CCCTGGTGCA ACTACACCAC 1020
 CGGTACAGTC GCCGTAATGT GGTCAGTACA GGGAGTGGCA TGGCTGAGGC CAGCCTGGAA 1080
 CCCAGCATAC GCTCGGGTAG CCAGCTCCTG GAGATGTTCC ACATTGGGCA GCAGCAGATC 1140
 5 TTTAAGCCCA CAGAGGATGA GGAAGAGAGT GAGGCCAAGT ACATTGGCTC AGCTGACTTC 1200
 CAGGCCAAGG AGATATTTAG CACCTGCCTG GAGGGAGAGC AGGGGCCACA GTTTGCGCCC 1260
 TCTGCCCCAC CCCTGAGCAC AGTGGACTCT GTATCCCAGG TGGCACCGGC AGCCCCTGTG 1320
 GAACCTGAAA CATTCCCTGA TAAGTATTCC CTGCAGTTTG GCTTTGGGCC TTTTGAGTTG 1380
 CCTCCTCAGT GGCTCTCAGA GACCCGAAAC AGCAAGAAGC GGCTGCTTCC CCCCTTGGGC 1440
 10 AACACCCCAG AAGAGCTGAT CCAGACAAAG GTGCCCAAGG TAGGCAGGGT GGAGCGGAAG 1500
 ATGAGCAGAA ACAATAAAGT GAGCATTTTT CCAAAGGTGG ATTCCTAG 1548

(221) INFORMATION FOR SEQ ID NO:220:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 515 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:220:

20 Met Gly His Asn Gly Ser Trp Ile Ser Pro Asn Ala Ser Glu Pro His
 1 5 10 15
 Asn Ala Ser Gly Ala Glu Ala Ala Gly Val Asn Arg Ser Ala Leu Gly
 20 25 30
 25 Glu Phe Gly Glu Ala Gln Leu Tyr Arg Gln Phe Thr Thr Thr Val Gln
 35 40 45
 Val Val Ile Phe Ile Gly Ser Leu Leu Gly Asn Phe Met Val Leu Trp
 50 55 60
 Ser Thr Cys Arg Thr Thr Val Phe Lys Ser Val Thr Asn Arg Phe Ile
 65 70 75 80
 30 Lys Asn Leu Ala Cys Ser Gly Ile Cys Ala Ser Leu Val Cys Val Pro
 85 90 95
 Phe Asp Ile Ile Leu Ser Thr Ser Pro His Cys Cys Trp Trp Ile Tyr
 100 105 110

	Thr Met Leu Phe Cys Lys Val Val Lys Phe Leu His Lys Val Phe Cys	
	115	120 125
	Ser Val Thr Ile Leu Ser Phe Pro Ala Ile Ala Leu Asp Arg Tyr Tyr	
	130	135 140
5	Ser Val Leu Tyr Pro Leu Glu Arg Lys Ile Ser Asp Ala Lys Ser Arg	
	145	150 155 160
	Glu Leu Val Met Tyr Ile Trp Ala His Ala Val Val Ala Ser Val Pro	
		165 170 175
10	Val Phe Ala Val Thr Asn Val Ala Asp Ile Tyr Ala Thr Ser Thr Cys	
		180 185 190
	Thr Glu Val Trp Ser Asn Ser Leu Gly His Leu Val Tyr Val Leu Val	
		195 200 205
	Tyr Asn Ile Thr Thr Val Ile Val Pro Val Val Val Val Phe Leu Phe	
		210 215 220
15	Leu Ile Leu Ile Arg Arg Ala Leu Ser Ala Ser Gln Lys Lys Lys Val	
		225 230 235 240
	Ile Ile Ala Ala Leu Arg Thr Pro Gln Asn Thr Ile Ser Ile Pro Tyr	
		245 250 255
20	Ala Ser Gln Arg Glu Ala Glu Leu Lys Ala Thr Leu Leu Ser Met Val	
		260 265 270
	Met Val Phe Ile Leu Cys Ser Val Pro Tyr Ala Thr Leu Val Val Tyr	
		275 280 285
	Gln Thr Val Leu Asn Val Pro Asp Thr Ser Val Phe Leu Leu Leu Thr	
		290 295 300
25	Ala Val Trp Leu Pro Lys Val Ser Leu Leu Ala Asn Pro Val Leu Phe	
		305 310 315 320
	Leu Thr Val Asn Lys Ser Val Arg Lys Cys Leu Ile Gly Thr Leu Val	
		325 330 335
30	Gln Leu His His Arg Tyr Ser Arg Arg Asn Val Val Ser Thr Gly Ser	
		340 345 350
	Gly Met Ala Glu Ala Ser Leu Glu Pro Ser Ile Arg Ser Gly Ser Gln	
		355 360 365
	Leu Leu Glu Met Phe His Ile Gly Gln Gln Gln Ile Phe Lys Pro Thr	
		370 375 380
35	Glu Asp Glu Glu Glu Ser Glu Ala Lys Tyr Ile Gly Ser Ala Asp Phe	
		385 390 395 400
	Gln Ala Lys Glu Ile Phe Ser Thr Cys Leu Glu Gly Glu Gln Gly Pro	

185

405 410 415

Gln Phe Ala Pro Ser Ala Pro Pro Leu Ser Thr Val Asp Ser Val Ser
420 425 430

5 Gln Val Ala Pro Ala Ala Pro Val Glu Pro Glu Thr Phe Pro Asp Lys
435 440 445

Tyr Ser Leu Gln Phe Gly Phe Gly Pro Phe Glu Leu Pro Pro Gln Trp
450 455 460

Leu Ser Glu Thr Arg Asn Ser Lys Lys Arg Leu Leu Pro Pro Leu Gly
465 470 475 480

10 Asn Thr Pro Glu Glu Leu Ile Gln Thr Lys Val Pro Lys Val Gly Arg
485 490 495

Val Glu Arg Lys Met Ser Arg Asn Asn Lys Val Ser Ile Phe Pro Lys
500 505 510

15 Val Asp Ser
515

(222) INFORMATION FOR SEQ ID NO:221:

(i) SEQUENCE CHARACTERISTICS:

- 20 (A) LENGTH: 1164 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:221:

ATGAATCGGC ACCATCTGCA GGATCACTTT CTGGAAATAG ACAAGAAGAA CTGCTGTGTG 60

25 TTCCGAGATG ACTTCATTGC CAAGGTGTTG CCGCCGGTGT TGGGGCTGGA GTTTATCTTT 120

GGGCTTCTGG GCAATGGCCT TGCCCTGTGG ATTTTCTGTT TCCACCTCAA GTCCTGGAAA 180

TCCAGCCGGA TTTTCCTGTT CAACCTGGCA GTAGCTGACT TTCTACTGAT CATCTGCCTG 240

CCGTTTCGTGA TGGACTACTA TGTGCGGCGT TCAGACTGGA AGTTTGGGGA CATCCCTTGC 300

CGGCTGGTGC TCTTCATGTT TGCCATGAAC CGCCAGGGCA GCATCATCTT CCTCACGGTG 360

30 GTGGCGGTAG ACAGGTATTT CCGGGTGGTC CATCCCCACC ACGCCCTGAA CAAGATCTCC 420

AATTGGACAG CAGCCATCAT CTCTTGCCCTT CTGTGGGGCA TCACTGTTGG CCTAACAGTC 480

CACCTCCTGA AGAAGAAGTT GCTGATCCAG AATGGCCCTG CAAATGTGTG CATCAGCTTC 540

AGCATCTGCC ATACCTTCCG GTGGCACGAA GCTATGTTCC TCCTGGAGTT CCTCCTGCCC 600

CTGGGCATCA TCCTGTTCTG CTCAGCCAGA ATTATCTGGA GCCTGCGGCA GAGACAAATG 660
 GACCGGCATG CCAAGATCAA GAGAGCCAAA ACCTTCATCA TGGTGGTGGC CATCGTCTTT 720
 GTCATCTGCT TCCTTCCCAG CGTGGTTGTG CGGATCCGCA TCTTCTGGCT CCTGCACACT 780
 TCGGGCACGC AGAATTGTGA AGTGTACCGC TCGGTGGACC TGGCGTTCTT TATCACTCTC 840
 5 AGCTTCACCT ACATGAACAG CATGCTGGAC CCCGTGGTGT ACTACTTCTC CAGCCCATCC 900
 TTTCCCAACT TCTTCTCCAC TTTGATCAAC CGCTGCCTCC AGAGGAAGAT GACAGGTGAG 960
 CCAGATAATA ACCGCAGCAC GAGCGTCGAG CTCACAGGGG ACCCCAACAA AACCAGAGGC 1020
 GCTCCAGAGG CGTTAATGGC CAACTCCGGT GAGCCATGGA GCCCCTCTTA TCTGGGCCCA 1080
 ACCTCAAATA ACCATTCCAA GAAGGGACAT TGTCACCAAG AACCAGCATC TCTGGAGAAA 1140
 10 CAGTTGGGCT GTTGCATCGA GTAA 1164

(223) INFORMATION FOR SEQ ID NO:222:

(i) SEQUENCE CHARACTERISTICS:

- 15 (A) LENGTH: 387 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:222:

20 Met Asn Arg His His Leu Gln Asp His Phe Leu Glu Ile Asp Lys Lys
 1 5 10 15
 Asn Cys Cys Val Phe Arg Asp Asp Phe Ile Ala Lys Val Leu Pro Pro
 20 25 30
 Val Leu Gly Leu Glu Phe Ile Phe Gly Leu Leu Gly Asn Gly Leu Ala
 35 40 45
 25 Leu Trp Ile Phe Cys Phe His Leu Lys Ser Trp Lys Ser Ser Arg Ile
 50 55 60
 Phe Leu Phe Asn Leu Ala Val Ala Asp Phe Leu Leu Ile Ile Cys Leu
 65 70 75 80
 30 Pro Phe Val Met Asp Tyr Tyr Val Arg Arg Ser Asp Trp Lys Phe Gly
 85 90 95
 Asp Ile Pro Cys Arg Leu Val Leu Phe Met Phe Ala Met Asn Arg Gln
 100 105 110
 Gly Ser Ile Ile Phe Leu Thr Val Val Ala Val Asp Arg Tyr Phe Arg
 115 120 125

187

	Val	Val	His	Pro	His	His	Ala	Leu	Asn	Lys	Ile	Ser	Asn	Trp	Thr	Ala	
	130						135					140					
	Ala	Ile	Ile	Ser	Cys	Leu	Leu	Trp	Gly	Ile	Thr	Val	Gly	Leu	Thr	Val	
	145					150					155					160	
5	His	Leu	Leu	Lys	Lys	Lys	Leu	Leu	Ile	Gln	Asn	Gly	Pro	Ala	Asn	Val	
					165					170						175	
	Cys	Ile	Ser	Phe	Ser	Ile	Cys	His	Thr	Phe	Arg	Trp	His	Glu	Ala	Met	
				180						185					190		
10	Phe	Leu	Leu	Glu	Phe	Leu	Leu	Pro	Leu	Gly	Ile	Ile	Leu	Phe	Cys	Ser	
				195				200					205				
	Ala	Arg	Ile	Ile	Trp	Ser	Leu	Arg	Gln	Arg	Gln	Met	Asp	Arg	His	Ala	
	210						215					220					
	Lys	Ile	Lys	Arg	Ala	Lys	Thr	Phe	Ile	Met	Val	Val	Ala	Ile	Val	Phe	
	225					230				235						240	
15	Val	Ile	Cys	Phe	Leu	Pro	Ser	Val	Val	Val	Arg	Ile	Arg	Ile	Phe	Trp	
					245					250					255		
	Leu	Leu	His	Thr	Ser	Gly	Thr	Gln	Asn	Cys	Glu	Val	Tyr	Arg	Ser	Val	
				260					265					270			
20	Asp	Leu	Ala	Phe	Phe	Ile	Thr	Leu	Ser	Phe	Thr	Tyr	Met	Asn	Ser	Met	
				275				280					285				
	Leu	Asp	Pro	Val	Val	Tyr	Tyr	Phe	Ser	Ser	Pro	Ser	Phe	Pro	Asn	Phe	
				290			295				300						
	Phe	Ser	Thr	Leu	Ile	Asn	Arg	Cys	Leu	Gln	Arg	Lys	Met	Thr	Gly	Glu	
	305					310				315					320		
25	Pro	Asp	Asn	Asn	Arg	Ser	Thr	Ser	Val	Glu	Leu	Thr	Gly	Asp	Pro	Asn	
					325					330					335		
	Lys	Thr	Arg	Gly	Ala	Pro	Glu	Ala	Leu	Met	Ala	Asn	Ser	Gly	Glu	Pro	
				340				345					350				
30	Trp	Ser	Pro	Ser	Tyr	Leu	Gly	Pro	Thr	Ser	Asn	Asn	His	Ser	Lys	Lys	
				355				360					365				
	Gly	His	Cys	His	Gln	Glu	Pro	Ala	Ser	Leu	Glu	Lys	Gln	Leu	Gly	Cys	
				370			375				380						
	Cys	Ile	Glu														
	385																

35 (224) INFORMATION FOR SEQ ID NO:223:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1212 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: DNA (genomic)

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:223:

	ATGGCTTGCA	ATGGCAGTGC	GGCCAGGGGG	CACTTTGACC	CTGAGGACTT	GAACCTGACT	60
	GACGAGGCAC	TGAGACTCAA	GTACCTGGGG	CCCCAGCAGA	CAGAGCTGTT	CATGCCCATC	120
	TGTGCCACAT	ACCTGCTGAT	CTTCGTGGTG	GGCGCTGTGG	GCAATGGGCT	GACCTGTCTG	180
	GTCATCCTGC	GCCACAAGGC	CATGCGCACG	CCTACCAACT	ACTACCTCTT	CAGCCTGGCC	240
10	GTGTCGGACC	TGCTGGTGCT	GCTGGTGGGC	CTGCCCCTGG	AGCTCTATGA	GATGTGGCAC	300
	AACTACCCCT	TCCTGCTGGG	CGTTGGTGGC	TGCTATTTC	GCACGCTACT	GTTTGAGATG	360
	GTCTGCCTGG	CCTCAGTGCT	CAACGTCACT	GCCCTGAGCG	TGGAACGCTA	TGTGGCCGTG	420
	GTGCACCCAC	TCCAGGCCAG	GTCCATGGTG	ACGCGGGCCC	ATGTGCGCCG	AGTGCTTGGG	480
	GCCGTCTGGG	GTCTTGCCAT	GCTCTGCTCC	CTGCCCAACA	CCAGCCTGCA	CGGCATCCGG	540
15	CAGCTGCACG	TGCCCTGCCG	GGGCCCAGTG	CCAGACTCAG	CTGTTTGCAT	GCTGGTCCGC	600
	CCACGGGCCC	TCTACAACAT	GGTAGTGCAG	ACCACCGCGC	TGCTCTTCTT	CTGCCTGCCC	660
	ATGGCCATCA	TGAGCGTGCT	CTACCTGCTC	ATTGGGCTGC	GACTGCGGCG	GGAGAGGCTG	720
	CTGCTCATGC	AGGAGGCCAA	GGGCAGGGGC	TCTGCAGCAG	CCAGGTCCAG	ATACACCTGC	780
	AGGCTCCAGC	AGCACGATCG	GGGCCGAGGA	CAAGTGAAGA	AGATGCTGTT	TGTCCTGGTC	840
20	GTGGTGTTTG	GCATCTGCTG	GGCCCCGTTC	CACGCCGACC	GCGTCATGTG	GAGCGTCGTG	900
	TCACAGTGGA	CAGATGGCCT	GCACCTGGCC	TTCCAGCACG	TGCACGTCAT	CTCCGGCATC	960
	TTCTTCTACC	TGGGCTCGGC	GGCCAACCCC	GTGCTCTATA	GCCTCATGTC	CAGCCGCTTC	1020
	CGAGAGACCT	TCCAGGAGGC	CCTGTGCCTC	GGGGCCTGCT	GCCATCGCCT	CAGACCCCGC	1080
	CACAGCTCCC	ACAGCCTCAG	CAGGATGACC	ACAGGCAGCA	CCCTGTGTGA	TGTGGGCTCC	1140
25	CTGGGCAGCT	GGGTCCACCC	CCTGGCTGGG	AACGATGGCC	CAGAGGCGCA	GCAAGAGACC	1200
	GATCCATCCT	GA					1212

(225) INFORMATION FOR SEQ ID NO:224:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 403 amino acids
(B) TYPE: amino acid

30

189

(C) STRANDEDNESS:

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:224:

5	Met	Ala	Cys	Asn	Gly	Ser	Ala	Ala	Arg	Gly	His	Phe	Asp	Pro	Glu	Asp	1	5	10	15
	Leu	Asn	Leu	Thr	Asp	Glu	Ala	Leu	Arg	Leu	Lys	Tyr	Leu	Gly	Pro	Gln	20	25	30	
10	Gln	Thr	Glu	Leu	Phe	Met	Pro	Ile	Cys	Ala	Thr	Tyr	Leu	Leu	Ile	Phe	35	40	45	
	Val	Val	Gly	Ala	Val	Gly	Asn	Gly	Leu	Thr	Cys	Leu	Val	Ile	Leu	Arg	50	55	60	
	His	Lys	Ala	Met	Arg	Thr	Pro	Thr	Asn	Tyr	Tyr	Leu	Phe	Ser	Leu	Ala	65	70	75	80
15	Val	Ser	Asp	Leu	Leu	Val	Leu	Val	Gly	Leu	Pro	Leu	Glu	Leu	Tyr		85	90	95	
	Glu	Met	Trp	His	Asn	Tyr	Pro	Phe	Leu	Leu	Gly	Val	Gly	Gly	Cys	Tyr	100	105	110	
20	Phe	Arg	Thr	Leu	Leu	Phe	Glu	Met	Val	Cys	Leu	Ala	Ser	Val	Leu	Asn	115	120	125	
	Val	Thr	Ala	Leu	Ser	Val	Glu	Arg	Tyr	Val	Ala	Val	Val	His	Pro	Leu	130	135	140	
	Gln	Ala	Arg	Ser	Met	Val	Thr	Arg	Ala	His	Val	Arg	Arg	Val	Leu	Gly	145	150	155	160
25	Ala	Val	Trp	Gly	Leu	Ala	Met	Leu	Cys	Ser	Leu	Pro	Asn	Thr	Ser	Leu	165	170	175	
	His	Gly	Ile	Arg	Gln	Leu	His	Val	Pro	Cys	Arg	Gly	Pro	Val	Pro	Asp	180	185	190	
30	Ser	Ala	Val	Cys	Met	Leu	Val	Arg	Pro	Arg	Ala	Leu	Tyr	Asn	Met	Val	195	200	205	
	Val	Gln	Thr	Thr	Ala	Leu	Leu	Phe	Phe	Cys	Leu	Pro	Met	Ala	Ile	Met	210	215	220	
	Ser	Val	Leu	Tyr	Leu	Leu	Ile	Gly	Leu	Arg	Leu	Arg	Arg	Glu	Arg	Leu	225	230	235	240
35	Leu	Leu	Met	Gln	Glu	Ala	Lys	Gly	Arg	Gly	Ser	Ala	Ala	Ala	Arg	Ser	245	250	255	

20

25

- (ii) MOLECULE TYPE: DNA (genomic)

35

RUSKIN IN - Wn 100-20A-1-

191

GTCAGCGTGG TCATCTGGGC CAAGGAGCTG CTGACCAGCA TCTACTTCCT GATGCACGAG 480
 GAGGTCATCG AGGACGAGAA CCAGCACCGC GTGTGCTTTG AGCACTACCC CATCCAGGCA 540
 TGGCAGCGCG CCATCAACTA CTACCGCTTC CTGGTGGGCT TCCTCTTCCC CATCTGCCTG 600
 CTGCTGGCGT CCTACCAGGG CATCCTGCGC GCCGTGCGCC GGAGCCACGG CACCCAGAAG 660
 5 AGCCGCAAGG ACCAGATCAA GCGGCTGGTG CTCAGCACCG TGGTCATCTT CCTGGCCTGC 720
 TTCCTGCCCT ACCACGTGTT GCTGCTGGTG CGCAGCGTCT GGGAGGCCAG CTGCGACTTC 780
 GCCAAGGGCG TTTTCAACGC CTACCACTTC TCCCTCCTGC TCACCAGCTT CAACTGCGTC 840
 GCCGACCCCG TGCTCTACTG CTTCGTCAGC GAGACCACCC ACCGGGACCT GGCCCGCCTC 900
 CGCGGGGCCT GCCTGGCCTT CCTCACCTGC TCCAGGACCG GCCGGGCCAG GGAGGCCTAC 960
 10 CCGCTGGGTG CCCCCGAGGC CTCCGGGAAA AGCGGGGCC AGGGTGAGGA GCCCGAGCTG 1020
 TTGACCAAGC TCCACCCGGC CTTCCAGACC CCTAACTCGC CAGGGTCGGG CGGGTTCCCC 1080
 ACGGGCAGGT TGGCCTAG 1098

(227) INFORMATION FOR SEQ ID NO:226:

- (i) SEQUENCE CHARACTERISTICS:
 15 (A) LENGTH: 365 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

- 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:226:

Met Gly Asn Ile Thr Ala Asp Asn Ser Ser Met Ser Cys Thr Ile Asp
 1 5 10 15

His Thr Ile His Gln Thr Leu Ala Pro Val Val Tyr Val Thr Val Leu
 20 25 30

- 25 Val Val Gly Phe Pro Ala Asn Cys Leu Ser Leu Tyr Phe Gly Tyr Leu
 35 40 45

Gln Ile Lys Ala Arg Asn Glu Leu Gly Val Tyr Leu Cys Asn Leu Thr
 50 55 60

- 30 Val Ala Asp Leu Phe Tyr Ile Cys Ser Leu Pro Phe Trp Leu Gln Tyr
 65 70 75 80

Val Leu Gln His Asp Asn Trp Ser His Gly Asp Leu Ser Cys Gln Val
 85 90 95

Cys Gly Ile Leu Leu Tyr Glu Asn Ile Tyr Ile Ser Val Gly Phe Leu

	100		105		110
	Cys Cys Ile Ser Val Asp Arg Tyr Leu Ala Val Ala His Pro Phe Arg				
	115		120		125
5	Phe His Gln Phe Arg Thr Leu Lys Ala Ala Val Gly Val Ser Val Val				
	130		135		140
	Ile Trp Ala Lys Glu Leu Leu Thr Ser Ile Tyr Phe Leu Met His Glu				
	145		150		155
	Glu Val Ile Glu Asp Glu Asn Gln His Arg Val Cys Phe Glu His Tyr				
		165		170	175
10	Pro Ile Gln Ala Trp Gln Arg Ala Ile Asn Tyr Tyr Arg Phe Leu Val				
		180		185	190
	Gly Phe Leu Phe Pro Ile Cys Leu Leu Leu Ala Ser Tyr Gln Gly Ile				
		195		200	205
15	Leu Arg Ala Val Arg Arg Ser His Gly Thr Gln Lys Ser Arg Lys Asp				
		210		215	220
	Gln Ile Lys Arg Leu Val Leu Ser Thr Val Val Ile Phe Leu Ala Cys				
		225		230	235
	Phe Leu Pro Tyr His Val Leu Leu Leu Val Arg Ser Val Trp Glu Ala				
		245		250	255
20	Ser Cys Asp Phe Ala Lys Gly Val Phe Asn Ala Tyr His Phe Ser Leu				
		260		265	270
	Leu Leu Thr Ser Phe Asn Cys Val Ala Asp Pro Val Leu Tyr Cys Phe				
		275		280	285
25	Val Ser Glu Thr Thr His Arg Asp Leu Ala Arg Leu Arg Gly Ala Cys				
		290		295	300
	Leu Ala Phe Leu Thr Cys Ser Arg Thr Gly Arg Ala Arg Glu Ala Tyr				
		305		310	315
	Pro Leu Gly Ala Pro Glu Ala Ser Gly Lys Ser Gly Ala Gln Gly Glu				
		325		330	335
30	Glu Pro Glu Leu Leu Thr Lys Leu His Pro Ala Phe Gln Thr Pro Asn				
		340		345	350
	Ser Pro Gly Ser Gly Gly Phe Pro Thr Gly Arg Leu Ala				
		355		360	365

(228) INFORMATION FOR SEQ ID NO:227:

- 35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1416 base pairs
 (B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:227:

5 ATGGATATTC TTTGTGAAGA AAATACTTCT TTGAGCTCAA CTACGAACTC CCTAATGCAA 60
TTAAATGATG ACAACAGGCT CTACAGTAAT GACTTTAACT CCGGAGAAGC TAACACTTCT 120
GATGCATTTA ACTGGACAGT CGACTCTGAA AATCGAACCA ACCTTTCCTG TGAAGGGTGC 180
CTCTCACCGT CGTGTCTCTC CTTACTTCAT CTCCAGGAAA AAAACTGGTC TGCTTTACTG 240
ACAGCCGTAG TGATTATTCT AACTATTGCT GGAAACATAC TCGTCATCAT GGCAGTGTCC 300
10 CTAGAGAAAA AGCTGCAGAA TGCCACCAAC TATTTCTTGA TGTCACCTGC CATAGCTGAT 360
ATGCTGCTGG GTTTCCTTGT CATGCCCCGTG TCCATGTTAA CCATCCTGTA TGGGTACCGG 420
TGGCCTCTGC CGAGCAAGCT TTGTGCAGTC TGGATTACC TGGACGTGCT CTTCTCCACG 480
GCCTCCATCA TGCACCTCTG CGCCATCTCG CTGGACCGCT ACGTCGCCAT CCAGAATCCC 540
ATCCACCACA GCCGCTTCAA CTCCAGAACT AAGGCATTTT TGAAAATCAT TGCTGTTTGG 600
15 ACCATATCAG TAGGTATATC CATGCCAATA CCAGTCTTTG GGCTACAGGA CGATTGGAAG 660
GTCTTTAAGG AGGGGAGTTG CTTACTCGCC GATGATAACT TTGTCCTGAT CGGCTCTTTT 720
GTGTCATTTT TCATTCCCTT AACCATCATG GTGATCACCT ACTTTCTAAC TATCAAGTCA 780
CTCCAGAAAG AAGCTACTTT GTGTGTAAGT GATCTTGGCA CACGGGCCAA ATTAGCTTCT 840
TTCAGCTTCC TCCCTCAGAG TTCTTTGTCT TCAGAAAAGC TCTTCCAGCG GTCGATCCAT 900
20 AGGGAGCCAG GGTCTACAC AGGCAGGAGG ACTATGCAGT CCATCAGCAA TGAGCAAAAG 960
GCAAAGAAGG TGCTGGGCAT CGTCTTCTTC CTGTTTGTGG TGATGTGGTG CCCTTTCTTC 1020
ATCACAAACA TCATGGCCGT CATCTGCAA GAGTCCTGCA ATGAGGATGT CATTGGGGCC 1080
CTGCTCAATG TGTTTGTTTG GATCGGTTAT CTCTCTTCAG CAGTCAACCC ACTAGTCTAC 1140
ACACTGTTCA ACAAGACCTA TAGGTCAGCC TTTTCACGGT ATATTCAGTG TCAGTACAAG 1200
25 GAAAACAAAA AACCATTGCA GTTAATTTTA GTGAACACAA TACCGGCTTT GGCCTACAAG 1260
TCTAGCCAAC TTCAAATGGG ACAAAAAAAG AATTCAAAGC AAGATGCCAA GACAACAGAT 1320
AATGACTGCT CAATGGTTGC TCTAGGAAAG CAGTATTCTG AAGAGGCTTC TAAAGACAAT 1380
AGCGACGGAG TGAATGAAAA GGTGAGCTGT GTGTGA 1416

(229) INFORMATION FOR SEQ ID NO:228:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 470 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:228:

```

10  1  Met Asp Ile Leu Cys Glu Glu Asn Thr Ser Leu Ser Ser Thr Thr Asn
      5              10              15
    Ser Leu Met Gln Leu Asn Asp Asp Asn Arg Leu Tyr Ser Asn Asp Phe
      20              25              30

    Asn Ser Gly Glu Ala Asn Thr Ser Asp Ala Phe Asn Trp Thr Val Asp
      35              40              45

15  Ser Glu Asn Arg Thr Asn Leu Ser Cys Glu Gly Cys Leu Ser Pro Ser
      50              55              60

    Cys Leu Ser Leu Leu His Leu Gln Glu Lys Asn Trp Ser Ala Leu Leu
      65              70              75              80

20  Thr Ala Val Val Ile Ile Leu Thr Ile Ala Gly Asn Ile Leu Val Ile
      85              90              95

    Met Ala Val Ser Leu Glu Lys Lys Leu Gln Asn Ala Thr Asn Tyr Phe
      100             105             110

    Leu Met Ser Leu Ala Ile Ala Asp Met Leu Leu Gly Phe Leu Val Met
      115             120             125

25  Pro Val Ser Met Leu Thr Ile Leu Tyr Gly Tyr Arg Trp Pro Leu Pro
      130             135             140

    Ser Lys Leu Cys Ala Val Trp Ile Tyr Leu Asp Val Leu Phe Ser Thr
      145             150             155             160

    Ala Ser Ile Met His Leu Cys Ala Ile Ser Leu Asp Arg Tyr Val Ala
      165             170             175

30  Ile Gln Asn Pro Ile His His Ser Arg Phe Asn Ser Arg Thr Lys Ala
      180             185             190

    Phe Leu Lys Ile Ile Ala Val Trp Thr Ile Ser Val Gly Ile Ser Met
      195             200             205

35  Pro Ile Pro Val Phe Gly Leu Gln Asp Asp Ser Lys Val Phe Lys Glu
      210             215             220

    Gly Ser Cys Leu Leu Ala Asp Asp Asn Phe Val Leu Ile Gly Ser Phe
      225             230             235             240
  
```

195

	Val	Ser	Phe	Phe	Ile	Pro	Leu	Thr	Ile	Met	Val	Ile	Thr	Tyr	Phe	Leu	
					245					250					255		
	Thr	Ile	Lys	Ser	Leu	Gln	Lys	Glu	Ala	Thr	Leu	Cys	Val	Ser	Asp	Leu	
				260					265					270			
5	Gly	Thr	Arg	Ala	Lys	Leu	Ala	Ser	Phe	Ser	Phe	Leu	Pro	Gln	Ser	Ser	
				275				280					285				
	Leu	Ser	Ser	Glu	Lys	Leu	Phe	Gln	Arg	Ser	Ile	His	Arg	Glu	Pro	Gly	
				290			295					300					
10	Ser	Tyr	Thr	Gly	Arg	Arg	Thr	Met	Gln	Ser	Ile	Ser	Asn	Glu	Gln	Lys	
	305					310					315					320	
	Ala	Lys	Lys	Val	Leu	Gly	Ile	Val	Phe	Phe	Leu	Phe	Val	Val	Met	Trp	
					325					330					335		
	Cys	Pro	Phe	Phe	Ile	Thr	Asn	Ile	Met	Ala	Val	Ile	Cys	Lys	Glu	Ser	
				340					345					350			
15	Cys	Asn	Glu	Asp	Val	Ile	Gly	Ala	Leu	Leu	Asn	Val	Phe	Val	Trp	Ile	
				355				360					365				
	Gly	Tyr	Leu	Ser	Ser	Ala	Val	Asn	Pro	Leu	Val	Tyr	Thr	Leu	Phe	Asn	
				370			375					380					
20	Lys	Thr	Tyr	Arg	Ser	Ala	Phe	Ser	Arg	Tyr	Ile	Gln	Cys	Gln	Tyr	Lys	
	385					390					395					400	
	Glu	Asn	Lys	Lys	Pro	Leu	Gln	Leu	Ile	Leu	Val	Asn	Thr	Ile	Pro	Ala	
					405					410					415		
	Leu	Ala	Tyr	Lys	Ser	Ser	Gln	Leu	Gln	Met	Gly	Gln	Lys	Lys	Asn	Ser	
				420					425					430			
25	Lys	Gln	Asp	Ala	Lys	Thr	Thr	Asp	Asn	Asp	Cys	Ser	Met	Val	Ala	Leu	
			435					440					445				
	Gly	Lys	Gln	Tyr	Ser	Glu	Glu	Ala	Ser	Lys	Asp	Asn	Ser	Asp	Gly	Val	
		450				455						460					
30	Asn	Glu	Lys	Val	Ser	Cys	Val										
	465					470											

(230) INFORMATION FOR SEQ ID NO:229:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1377 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:229:

ATGGTGAACC TGAGGAATGC GGTGCATTCA TTCCTTGTGC ACCTAATTGG CCTATTGGTT 60
 TGGCAATGTG ATATTCTGT GAGCCCAGTA GCAGCTATAG TAACTGACAT TTTCAATACC 120
 TCCGATGGTG GACGCTTCAA ATTCCCAGAC GGGGTACAAA ACTGGCCAGC ACTTTCAATC 180
 5 GTCATCATAA TAATCATGAC AATAGGTGGC AACATCCTTG TGATCATGGC AGTAAGCATG 240
 GAAAAGAAAC TGCACAATGC CACCAATTAC TTCTTAATGT CCCTAGCCAT TGCTGATATG 300
 CTAGTGGGAC TACTTGTCAT GCCCCTGTCT CTCCTGGCAA TCCTTTATGA TTATGTCTGG 360
 CCACTACCTA GATATTTGTG CCCCCTCTGG ATTTCTTTAG ATGTTTTATT TTCAACAGCG 420
 TCCATCATGC ACCTCTGCGC TATATCGCTG GATCGGTATG TAGCAATACG TAATCCTATT 480
 10 GAGCATAGCC GTTTC AATTC GCGGACTAAG GCCATCATGA AGATTGCTAT TGTTTGGGCA 540
 ATTTCTATAG GTGTATCAGT TCCTATCCCT GTGATTGGAC TGAGGGACGA AGAAAAGGTG 600
 TTCGTGAACA ACACGACGTG CGTGCTCAAC GACCCAAATT TCGTTCTTAT TGGGTCCTTC 660
 GTAGCTTTCT TCATACCGCT GACGATTATG GTGATTACGT ATTGCCTGAC CATCTACGTT 720
 CTGCGCCGAC AAGCTTTGAT GTTACTGCAC GGCCACACCG AGGAACCGCC TGGACTAAGT 780
 15 CTGGATTTCC TGAAGTGCTG CAAGAGGAAT ACGGCCGAGG AAGAGAACTC TGCAAACCCT 840
 AACCAAGACC AGAACGCACG CCGAAGAAAG AAGAAGGAGA GACGTCCTAG GGGCACCATG 900
 CAGGCTATCA ACAATGAAAG AAAAGCTAAG AAAGTCCTTG GGATTGTTTT CTTTGTGTTT 960
 CTGATCATGT GGTGCCCATT TTTCATTACC AATATTCTGT CTGTTCTTTG TGAGAAGTCC 1020
 TGTAACCAAA AGCTCATGGA AAAGCTTCTG AATGTGTTTG TTTGGATTGG CTATGTTTGT 1080
 20 TCAGGAATCA ATCCTCTGGT GTATACTCTG TTCAACAAAA TTTACCGAAG GGCATTCTCC 1140
 AACTATTTGC GTTGCAATTA TAAGGTAGAG AAAAGCCTC CTGTCAGGCA GATTCCAAGA 1200
 GTTGCCGCCA CTGCTTTGTC TGGGAGGGAG CTTAATGTTA ACATTTATCG GCATACCAAT 1260
 GAACCGGTGA TCGAGAAAGC CAGTGACAAT GAGCCCGGTA TAGAGATGCA AGTTGAGAAAT 1320
 TTAGAGTTAC CAGTAAATCC CTCCAGTGTG GTTAGCGAAA GGATTAGCAG TGTGTGA 1377

25 (231) INFORMATION FOR SEQ ID NO:230:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 458 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

197

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:230:

```

5      Met Val Asn Leu Arg Asn Ala Val His Ser Phe Leu Val His Leu Ile
      1      5      10      15

      Gly Leu Leu Val Trp Gln Cys Asp Ile Ser Val Ser Pro Val Ala Ala
      20      25      30

      Ile Val Thr Asp Ile Phe Asn Thr Ser Asp Gly Gly Arg Phe Lys Phe
      35      40      45

10     Pro Asp Gly Val Gln Asn Trp Pro Ala Leu Ser Ile Val Ile Ile Ile
      50      55      60

      Ile Met Thr Ile Gly Gly Asn Ile Leu Val Ile Met Ala Val Ser Met
      65      70      75      80

15     Glu Lys Lys Leu His Asn Ala Thr Asn Tyr Phe Leu Met Ser Leu Ala
      85      90      95

      Ile Ala Asp Met Leu Val Gly Leu Leu Val Met Pro Leu Ser Leu Leu
      100     105     110

      Ala Ile Leu Tyr Asp Tyr Val Trp Pro Leu Pro Arg Tyr Leu Cys Pro
      115     120     125

20     Val Trp Ile Ser Leu Asp Val Leu Phe Ser Thr Ala Ser Ile Met His
      130     135     140

      Leu Cys Ala Ile Ser Leu Asp Arg Tyr Val Ala Ile Arg Asn Pro Ile
      145     150     155     160

25     Glu His Ser Arg Phe Asn Ser Arg Thr Lys Ala Ile Met Lys Ile Ala
      165     170     175

      Ile Val Trp Ala Ile Ser Ile Gly Val Ser Val Pro Ile Pro Val Ile
      180     185     190

      Gly Leu Arg Asp Glu Glu Lys Val Phe Val Asn Asn Thr Thr Cys Val
      195     200     205

30     Leu Asn Asp Pro Asn Phe Val Leu Ile Gly Ser Phe Val Ala Phe Phe
      210     215     220

      Ile Pro Leu Thr Ile Met Val Ile Thr Tyr Cys Leu Thr Ile Tyr Val
      225     230     235     240

35     Leu Arg Arg Gln Ala Leu Met Leu Leu His Gly His Thr Glu Glu Pro
      245     250     255

      Pro Gly Leu Ser Leu Asp Phe Leu Lys Cys Cys Lys Arg Asn Thr Ala

```

	260	265	270
	Glu Glu Glu Asn Ser Ala Asn Pro Asn Gln Asp Gln Asn Ala Arg Arg		
	275	280	285
5	Arg Lys Lys Lys Glu Arg Arg Pro Arg Gly Thr Met Gln Ala Ile Asn		
	290	295	300
	Asn Glu Arg Lys Ala Lys Lys Val Leu Gly Ile Val Phe Phe Val Phe		
	305	310	315 320
	Leu Ile Met Trp Cys Pro Phe Phe Ile Thr Asn Ile Leu Ser Val Leu		
	325	330	335
10	Cys Glu Lys Ser Cys Asn Gln Lys Leu Met Glu Lys Leu Leu Asn Val		
	340	345	350
	Phe Val Trp Ile Gly Tyr Val Cys Ser Gly Ile Asn Pro Leu Val Tyr		
	355	360	365
15	Thr Leu Phe Asn Lys Ile Tyr Arg Arg Ala Phe Ser Asn Tyr Leu Arg		
	370	375	380
	Cys Asn Tyr Lys Val Glu Lys Lys Pro Pro Val Arg Gln Ile Pro Arg		
	385	390	395 400
	Val Ala Ala Thr Ala Leu Ser Gly Arg Glu Leu Asn Val Asn Ile Tyr		
	405	410	415
20	Arg His Thr Asn Glu Pro Val Ile Glu Lys Ala Ser Asp Asn Glu Pro		
	420	425	430
	Gly Ile Glu Met Gln Val Glu Asn Leu Glu Leu Pro Val Asn Pro Ser		
	435	440	445
25	Ser Val Val Ser Glu Arg Ile Ser Ser Val		
	450	455	

(232) INFORMATION FOR SEQ ID NO:231:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1068 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:231:

	ATGGATCAGT TCCCTGAATC AGTGACAGAA AACTTTGAGT ACGATGATTT GGCTGAGGCC	60
35	TGTTATATTG GGGACATCGT GGTCTTTGGG ACTGTGTTCC TGTCCATATT CTACTCCGTC	120
	ATCTTTGCCA TTGGCCTGGT GGGAAATTG TTGGTAGTGT TTGCCCTCAC CAACAGCAAG	180

199

AAGCCCAAGA GTGTCACCGA CATTACCTC CTGAACCTGG CCTTGTCTGA TCTGCTGTTT 240
 GTAGCCACTT TGCCCTTCTG GACTCACTAT TTGATAAATG AAAAGGGCCT CCACAATGCC 300
 ATGTGCAAAT TCACTACCGC CTTCTTCTTC ATCGGCTTTT TTGGAAGCAT ATTCTTCATC 360
 ACCGTCATCA GCATTGATAG GTACCTGGCC ATCGTCCTGG CCGCCAACTC CATGAACAAC 420
 5 CGGACCGTGC AGCATGGCGT CACCATCAGC CTAGGCGTCT GGGCAGCAGC CATTTTGGTG 480
 GCAGCACCCC AGTTCATGTT CACAAAGCAG AAAGAAAATG AATGCCTTGG TGACTIONCCC 540
 GAGGTCCTCC AGGAAATCTG GCGCGTGCTC CGCAATGTGG AAACAAATTT TCTTGGCTTC 600
 CTACTCCCCC TGCTCATTAT GAGTTATTGC TACTTCAGAA TCATCCAGAC GCTGTTTTCC 660
 TGCAAGAACC ACAAGAAAGC CAAAGCCAAG AAAGTATCC TTCTGGTGGT CATCGTGTTT 720
 10 TTCCTCTTCT GGACACCCTA CAACGTTATG ATTTTCCTGG AGACGCTTAA GCTCTATGAC 780
 TTCTTTCCCA GTTGTGACAT GAGGAAGGAT CTGAGGCTGG CCCTCAGTGT GACTGAGACG 840
 GTTGCAATTA GCCATTGTTG CCTGAATCCT CTCATCTATG CATTGCTGG GGAGAAGTTC 900
 AGAAGATACC TTTACCACCT GTATGGGAAA TGCCTGGCTG TCCTGTGTGG GCGCTCAGTC 960
 CACGTTGATT TCTCCTCATC TGAATCACAA AGGAGCAGGC ATGGAAGTGT TCTGAGCAGC 1020
 15 AATTTTACTT ACCACACGAG TGATGGAGAT GCATTGCTCC TTCTCTGA 1068

(233) INFORMATION FOR SEQ ID NO:232:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 355 amino acids
 (B) TYPE: amino acid
 20 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:232:

25 Met Asp Gln Phe Pro Glu Ser Val Thr Glu Asn Phe Glu Tyr Asp Asp
 1 5 10 15
 Leu Ala Glu Ala Cys Tyr Ile Gly Asp Ile Val Val Phe Gly Thr Val
 20 25 30
 Phe Leu Ser Ile Phe Tyr Ser Val Ile Phe Ala Ile Gly Leu Val Gly
 35 40 45
 30 Asn Leu Leu Val Val Phe Ala Leu Thr Asn Ser Lys Lys Pro Lys Ser
 50 55 60
 Val Thr Asp Ile Tyr Leu Leu Asn Leu Ala Leu Ser Asp Leu Leu Phe

	65		70		75		80									
	Val	Ala	Thr	Leu	Pro	Phe	Trp	Thr	His	Tyr	Leu	Ile	Asn	Glu	Lys	Gly
				85						90					95	
5	Leu	His	Asn	Ala	Met	Cys	Lys	Phe	Thr	Thr	Ala	Phe	Phe	Phe	Ile	Gly
				100					105						110	
	Phe	Phe	Gly	Ser	Ile	Phe	Phe	Ile	Thr	Val	Ile	Ser	Ile	Asp	Arg	Tyr
			115						120					125		
	Leu	Ala	Ile	Val	Leu	Ala	Ala	Asn	Ser	Met	Asn	Asn	Arg	Thr	Val	Gln
		130					135						140			
10	His	Gly	Val	Thr	Ile	Ser	Leu	Gly	Val	Trp	Ala	Ala	Ala	Ile	Leu	Val
	145					150				155					160	
	Ala	Ala	Pro	Gln	Phe	Met	Phe	Thr	Lys	Gln	Lys	Glu	Asn	Glu	Cys	Leu
				165					170						175	
15	Gly	Asp	Tyr	Pro	Glu	Val	Leu	Gln	Glu	Ile	Trp	Pro	Val	Leu	Arg	Asn
			180					185						190		
	Val	Glu	Thr	Asn	Phe	Leu	Gly	Phe	Leu	Leu	Pro	Leu	Leu	Ile	Met	Ser
		195					200							205		
	Tyr	Cys	Tyr	Phe	Arg	Ile	Ile	Gln	Thr	Leu	Phe	Ser	Cys	Lys	Asn	His
	210					215						220				
20	Lys	Lys	Ala	Lys	Ala	Lys	Lys	Leu	Ile	Leu	Leu	Val	Val	Ile	Val	Phe
	225					230				235					240	
	Phe	Leu	Phe	Trp	Thr	Pro	Tyr	Asn	Val	Met	Ile	Phe	Leu	Glu	Thr	Leu
				245					250					255		
25	Lys	Leu	Tyr	Asp	Phe	Phe	Pro	Ser	Cys	Asp	Met	Arg	Lys	Asp	Leu	Arg
			260					265						270		
	Leu	Ala	Leu	Ser	Val	Thr	Glu	Thr	Val	Ala	Phe	Ser	His	Cys	Cys	Leu
		275					280						285			
	Asn	Pro	Leu	Ile	Tyr	Ala	Phe	Ala	Gly	Glu	Lys	Phe	Arg	Arg	Tyr	Leu
	290					295						300				
30	Tyr	His	Leu	Tyr	Gly	Lys	Cys	Leu	Ala	Val	Leu	Cys	Gly	Arg	Ser	Val
	305				310					315					320	
	His	Val	Asp	Phe	Ser	Ser	Ser	Glu	Ser	Gln	Arg	Ser	Arg	His	Gly	Ser
				325					330					335		
35	Val	Leu	Ser	Ser	Asn	Phe	Thr	Tyr	His	Thr	Ser	Asp	Gly	Asp	Ala	Leu
			340					345					350			
	Leu	Leu	Leu													
			355													

201

(234) INFORMATION FOR SEQ ID NO:233:

- 5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:233:

10 GGCTTAAGAG CATCATCGTG GTGCTGGTG

29

(235) INFORMATION FOR SEQ ID NO:234:

- 15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:234:

20 GTCACCACCA GCACCACGAT GATGCTCTTA AGCC

34

(236) INFORMATION FOR SEQ ID NO:235:

- 25 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:235:

CAAAGAAAGT ACTGGGCATC GTCTTCTTCC T

31

30 (237) INFORMATION FOR SEQ ID NO:236:

- 35 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:236:

TGCTCTAGAT TCCAGATAGG TGAAAACCTTG

30

(238) INFORMATION FOR SEQ ID NO:237:

- 5 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 50 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- 10 (iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:237:

CTAGGGGCAC CATGCAGGCT ATCAACAATG AAAGAAAAGC TAAGAAAGTC

50

(239) INFORMATION FOR SEQ ID NO:238:

- 15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 50 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- 20 (iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:238:

CAAGGACTTT CTTAGCTTTT CTTTCATTGT TGATAGCCTG CATGGTGCCC

50

(240) INFORMATION FOR SEQ ID NO:239:

- 25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 35 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:239:

CGGCGGCAGA AGGCGAAACG CATGATCCTC GCGGT

35

(241) INFORMATION FOR SEQ ID NO:240:

- 35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 35 base pairs
 (B) TYPE: nucleic acid

203

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:240:

5 ACCGCGAGGA TCATGCGTTT CGCCTTCTGC CGCCG 35

(242) INFORMATION FOR SEQ ID NO:241:

(i) SEQUENCE CHARACTERISTICS:

10 (A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:241:

GAGACATATT ATCTGCCACG GAGG 24

15 (243) INFORMATION FOR SEQ ID NO:242:

(i) SEQUENCE CHARACTERISTICS:

20 (A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:242:

TTGGCATAGA AACCGGACCC AAGG 24

(244) INFORMATION FOR SEQ ID NO:243:

25 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

30 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:243:

TAAGAATTCC ATAAAAATTA TGGAATGG 28

(245) INFORMATION FOR SEQ ID NO:244:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:244:

CCAGGATCCA GCTGAAGTCT TCCATCATTC

30

(246) INFORMATION FOR SEQ ID NO:245:

- 10 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1071 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:245:

ATGAATGGGG TCTCGGAGGG GACCAGAGGC TGCAGTGACA GGCAACCTGG GGTCTGACA 60

CGTGATCGCT CTTGTTCCAG GAAGATGAAC TCTTCCGGAT GCCTGTCTGA GGAGGTGGGG 120

TCCCTCCGCC CACTGACTGT GGTATCCTG TCTGCGTCCA TTGTCGTCGG AGTGCTGGGC 180

AATGGGCTGG TGCTGTGGAT GACTGTCTTC CGTATGGCAC GCACGGTCTC CACCGTCTGC 240

20 TTCTTCCACC TGGCCCTTGC CGATTTTCATG CTCTCACTGT CTCTGCCCCAT TGCCATGTAC 300

TATATTGTCT CCAGGCAGTG GTCCTCGGA GAGTGGGCCT GCAAACCTCTA CATCACCTTT 360

GTGTTCCCTCA GCTACTTTGC CAGTAACTGC CTCCTTGTCT TCATCTCTGT GGACCGTTGC 420

ATCTCTGTCC TCTACCCCGT CTGGGCCCTG AACCACCGCA CTGTGCAGCG GGCGAGCTGG 480

CTGGCCTTTG GGGTGTGGCT CCTGGCCGCC GCCTTGTGCT CTGCGCACCT GAAATTCCGG 540

25 ACAACCAGAA AATGGAATGG CTGTACGCAC TGCTACTTGG CGTTCAACTC TGACAATGAG 600

ACTGCCCAGA TTTGGATTGA AGGGGTCGTG GAGGGACACA TTATAGGGAC CATTGGCCAC 660

TTCCTGCTGG GCTTCCTGGG GCCCTTAGCA ATCATAGGCA CCTGCGCCCA CCTCATCCGG 720

GCCAAGCTCT TCGGGGAGGG CTGGGTCCAT GCCAACCGGC CCGCGAGGCT GCTGCTGGTG 780

CTGGTGAGCG CTTTCTTTAT CTTCTGGTCC CCGTTTAACG TGGTGCTGTT GGTCCATCTG 840

30 TGGCGACGGG TGATGCTCAA GGAAATCTAC CACCCCCGGA TGCTGCTCAT CCTCCAGGCT 900

AGCTTTGCCT TGGGCTGTGT CAACAGCAGC CTCAACCCCT TCCTCTACGT CTTCTGTTGGC 960

205

AGAGATTTCC AAGAAAAGTT TTTCCAGTCT TTGACTTCTG CCCTGGCGAG GGCCTTTGGA 1020
 GAGGAGGAGT TTCTGTCATC CTGTCCCCGT GGCAACGCCC CCCGGAATG A 1071

(247) INFORMATION FOR SEQ ID NO:246:

- 5 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 356 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

- 10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:246:

Met	Asn	Gly	Val	Ser	Glu	Gly	Thr	Arg	Gly	Cys	Ser	Asp	Arg	Gln	Pro	1	5	10	15
Gly	Val	Leu	Thr	Arg	Asp	Arg	Ser	Cys	Ser	Arg	Lys	Met	Asn	Ser	Ser	20	25	30	
Gly	Cys	Leu	Ser	Glu	Glu	Val	Gly	Ser	Leu	Arg	Pro	Leu	Thr	Val	Val	35	40	45	
Ile	Leu	Ser	Ala	Ser	Ile	Val	Val	Gly	Val	Leu	Gly	Asn	Gly	Leu	Val	50	55	60	
Leu	Trp	Met	Thr	Val	Phe	Arg	Met	Ala	Arg	Thr	Val	Ser	Thr	Val	Cys	65	70	75	80
Phe	Phe	His	Leu	Ala	Leu	Ala	Asp	Phe	Met	Leu	Ser	Leu	Ser	Leu	Pro	85	90	95	
Ile	Ala	Met	Tyr	Tyr	Ile	Val	Ser	Arg	Gln	Trp	Leu	Leu	Gly	Glu	Trp	100	105	110	
Ala	Cys	Lys	Leu	Tyr	Ile	Thr	Phe	Val	Phe	Leu	Ser	Tyr	Phe	Ala	Ser	115	120	125	
Asn	Cys	Leu	Leu	Val	Phe	Ile	Ser	Val	Asp	Arg	Cys	Ile	Ser	Val	Leu	130	135	140	
Tyr	Pro	Val	Trp	Ala	Leu	Asn	His	Arg	Thr	Val	Gln	Arg	Ala	Ser	Trp	145	150	155	160
Leu	Ala	Phe	Gly	Val	Trp	Leu	Leu	Ala	Ala	Ala	Leu	Cys	Ser	Ala	His	165	170	175	
Leu	Lys	Phe	Arg	Thr	Thr	Arg	Lys	Trp	Asn	Gly	Cys	Thr	His	Cys	Tyr	180	185	190	
Leu	Ala	Phe	Asn	Ser	Asp	Asn	Glu	Thr	Ala	Gln	Ile	Trp	Ile	Glu	Gly	195	200	205	

Val Val Glu Gly His Ile Ile Gly Thr Ile Gly His Phe Leu Leu Gly
 210 215 220

Phe Leu Gly Pro Leu Ala Ile Ile Gly Thr Cys Ala His Leu Ile Arg
 225 230 235 240

5 Ala Lys Leu Leu Arg Glu Gly Trp Val His Ala Asn Arg Pro Ala Arg
 245 250 255

Leu Leu Leu Val Leu Val Ser Ala Phe Phe Ile Phe Trp Ser Pro Phe
 260 265 270

10 Asn Val Val Leu Leu Val His Leu Trp Arg Arg Val Met Leu Lys Glu
 275 280 285

Ile Tyr His Pro Arg Met Leu Leu Ile Leu Gln Ala Ser Phe Ala Leu
 290 295 300

Gly Cys Val Asp Ser Ser Leu Asn Pro Phe Leu Tyr Val Phe Val Gly
 305 310 315 320

15 Arg Asp Phe Gln Glu Lys Phe Phe Gln Ser Leu Thr Ser Ala Leu Ala
 325 330 335

Arg Ala Phe Gly Glu Glu Glu Phe Leu Ser Ser Cys Pro Arg Gly Asn
 340 345 350

20 Ala Pro Arg Glu
 355

(248) INFORMATION FOR SEQ ID NO:247:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 32 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:247:

GCAGAATTTCG GCGGCCCCAT GGACCTGCCC CC

32

30 (249) INFORMATION FOR SEQ ID NO:248:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:248:

35

207

GCTGGATCCC CCGAGCAGTG GCGTTACTTC

30

(250) INFORMATION FOR SEQ ID NO:249:

(i) SEQUENCE CHARACTERISTICS:

5

(A) LENGTH: 903 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:249:

10 ATGGACCTGC CCCCAGCAGCT CTCCTTCGGC CTCTATGTGG CCGCCTTTGC GCTGGGCTTC 60
CCGCTCAACG TCCTGGCCAT CCGAGGCGCG ACGGCCACG CCCGGCTCCG TCTACCCCCT 120
AGCCTGGTCT ACGCCCTGAA CCTGGGCTGC TCCGACCTGC TGCTGACAGT CTCTCTGCCC 180
CTGAAGGCGG TGGAGGCGCT AGCCTCCGGG GCCTGGCCTC TGCCGGCCTC GCTGTGCCCC 240
GTCTTCGCGG TGGCCCACTT CTTCCCACTC TATGCCGGCG GGGGCTTCCT GGCCGCCCTG 300
15 AGTGCAGGCC GCTACCTGGG AGCAGCCTTC CCCTTGGGCT ACCAAGCCTT CCGGAGGCCG 360
TGCTATTCTT GGGGGGTGTG CGCGGCCATC TGGGCCCTCG TCCTGTGTCA CCTGGGTCTG 420
GTCTTTGGGT TGGAGGCTCC AGGAGGCTGG CTGGACCACA GCAACACCTC CCTGGGCATC 480
AACACACCGG TCAACGGCTC TCCGGTCTGC CTGGAGGCCT GGGACCCGGC CTCTGCCGGC 540
CCGGCCCGCT TCAGCCTCTC TCTCCTGCTC TTTTCTGCTC CTTGGCCAT CACAGCCTTC 600
20 TGCTACGTGG GCTGCCTCCG GGCAGTGGCC CGCTCCGGCC TGACGCACAG GCGGAAGCTG 660
CGGGCCGCCT GGGTGGCCGG CGGGGCCCTC CTCACGCTGC TGCTCTGCGT AGGACCCTAC 720
AACGCCTCCA ACGTGGCCAG CTTCTGTAC CCCAATCTAG GAGGCTCCTG GCGGAAGCTG 780
GGGCTCATCA CGGGTGCTG GAGTGTGGTG CTTAATCCGC TGGTGACCGG TTAATTGGGA 840
AGGGGTCCTG GCCTGAAGAC AGTGTGTGCG GCAAGAACGC AAGGGGGCAA GTCCCAGAAG 900
25 TAA 903

(251) INFORMATION FOR SEQ ID NO:250:

(i) SEQUENCE CHARACTERISTICS:

30

(A) LENGTH: 300 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:250:

Met Asp Leu Pro Pro Gln Leu Ser Phe Gly Leu Tyr Val Ala Ala Phe
 1 5 10 15
 5 Ala Leu Gly Phe Pro Leu Asn Val Leu Ala Ile Arg Gly Ala Thr Ala
 20 25 30
 His Ala Arg Leu Arg Leu Thr Pro Ser Leu Val Tyr Ala Leu Asn Leu
 35 40 45
 10 Gly Cys Ser Asp Leu Leu Leu Thr Val Ser Leu Pro Leu Lys Ala Val
 50 55 60
 Glu Ala Leu Ala Ser Gly Ala Trp Pro Leu Pro Ala Ser Leu Cys Pro
 65 70 75 80
 Val Phe Ala Val Ala His Phe Phe Pro Leu Tyr Ala Gly Gly Gly Phe
 85 90 95
 15 Leu Ala Ala Leu Ser Ala Gly Arg Tyr Leu Gly Ala Ala Phe Pro Leu
 100 105 110
 Gly Tyr Gln Ala Phe Arg Arg Pro Cys Tyr Ser Trp Gly Val Cys Ala
 115 120 125
 20 Ala Ile Trp Ala Leu Val Leu Cys His Leu Gly Leu Val Phe Gly Leu
 130 135 140
 Glu Ala Pro Gly Gly Trp Leu Asp His Ser Asn Thr Ser Leu Gly Ile
 145 150 155 160
 Asn Thr Pro Val Asn Gly Ser Pro Val Cys Leu Glu Ala Trp Asp Pro
 165 170 175
 25 Ala Ser Ala Gly Pro Ala Arg Phe Ser Leu Ser Leu Leu Phe Phe
 180 185 190
 Leu Pro Leu Ala Ile Thr Ala Phe Cys Tyr Val Gly Cys Leu Arg Ala
 195 200 205
 30 Leu Ala Arg Ser Gly Leu Thr His Arg Arg Lys Leu Arg Ala Ala Trp
 210 215 220
 Val Ala Gly Gly Ala Leu Leu Thr Leu Leu Leu Cys Val Gly Pro Tyr
 225 230 235 240
 Asn Ala Ser Asn Val Ala Ser Phe Leu Tyr Pro Asn Leu Gly Gly Ser
 245 250 255
 35 Trp Arg Lys Leu Gly Leu Ile Thr Gly Ala Trp Ser Val Val Leu Asn
 260 265 270
 Pro Leu Val Thr Gly Tyr Leu Gly Arg Gly Pro Gly Leu Lys Thr Val

209

275

280

285

Cys Ala Ala Arg Thr Gln Gly Gly Lys Ser Gln Lys
 290 295 300

(252) INFORMATION FOR SEQ ID NO:251:

- 5 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 31 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- 10 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:251:

CTCAAGCTTA CTCTCTCTCA CCAGTGGCCA C

31

(253) INFORMATION FOR SEQ ID NO:252:

- 15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 24 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:252:

CCCTCCTCCC CCGGAGGACC TAGC

24

(254) INFORMATION FOR SEQ ID NO:253:

- 25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1041 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:253:

30 ATGGATACAG GCCCCGACCA GTCCTACTTC TCCGGCAATC ACTGGTTCGT CTTCTCGGTG 60
 TACCTTCTCA CTTTCCTGGT GGGGCTCCCC CTCAACCTGC TGGCCCTGGT GGTCTTCGTG 120
 GGCAAGCTGC AGCGCCGCCC GGTGGCCGTG GACGTGCTCC TGCTCAACCT GACCGCCTCG 180
 GACCTGCTCC TGCTGCTGTT CCTGCCTTTC CGCATGGTGG AGGCAGCCAA TGGCATGCAC 240
 TGGCCCCCTGC CCTTCATCCT CTGCCCACTC TCTGGATTCA TCTTCTTCAC CACCATCTAT 300

CTCACCGCCC TCTTCCTGGC AGCTGTGAGC ATTGAACGCT TCCTGAGTGT GGCCACCCCA 360
 CTGTGGTACA AGACCCGGCC GAGGCTGGGG CAGGCAGGTC TGGTGAGTGT GGCCTGCTGG 420
 CTGTTGGCCT CTGCTCACTG CAGCGTGGTC TACGTCATAG AATTCTCAGG GGACATCTCC 480
 CACAGCCAGG GCACCAATGG GACCTGCTAC CTGGAGTTCC GGAAGGACCA GCTAGCCATC 540
 5 CTCCTGCCCCG TCGGGCTGGA GATGGCTGTG GTCCTCTTTG TGGTCCCGCT GATCATCACC 600
 AGCTACTGCT ACAGCCGCCT GGTGTGGATC CTCGGCAGAG GGGGCAGCCA CCGCCGGCAG 660
 AGGAGGGTGG CGGGGCTGTT GCGGGCCACG CTGCTCAACT TCCTGTCTG CTTTGGGCCC 720
 TACAACGTGT CCCATGTCGT GGGCTATATC TCGGTGAAA GCCCGGCATG GAGGATCTAC 780
 GTGACGCTTC TCAGCACCTT GAACTCCTGT GTCGACCCCT TTGTCTACTA CTTCTCCTCC 840
 10 TCCGGGTTC AAGCCGACTT TCATGAGCTG CTGAGGAGGT TGTGTGGGCT CTGGGGCCAG 900
 TGGCAGCAGG AGAGCAGCAT GGAGCTGAAG GAGCAGAAGG GAGGGGAGGA GCAGAGAGCG 960
 GACCGACCAG CTGAAAGAAA GACCACTGAA CACTCACAGG GCTGTGGAAC TGGTGGCCAG 1020
 GTGCCTGTG CTGAAAGCTA G 1041

(255) INFORMATION FOR SEQ ID NO:254:

- 15 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 346 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
- 20 (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:254:
- | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Asp | Thr | Gly | Pro | Asp | Gln | Ser | Tyr | Phe | Ser | Gly | Asn | His | Trp | Phe |
| 1 | | | | 5 | | | | | 10 | | | | | 15 | |
| Val | Phe | Ser | Val | Tyr | Leu | Leu | Thr | Phe | Leu | Val | Gly | Leu | Pro | Leu | Asn |
| 25 | | | 20 | | | | | 25 | | | | | 30 | | |
| Leu | Leu | Ala | Leu | Val | Val | Phe | Val | Gly | Lys | Leu | Gln | Arg | Arg | Pro | Val |
| | | 35 | | | | 40 | | | | | | 45 | | | |
| Ala | Val | Asp | Val | Leu | Leu | Leu | Asn | Leu | Thr | Ala | Ser | Asp | Leu | Leu | Leu |
| | | 50 | | | | 55 | | | | | 60 | | | | |
| Leu | Leu | Phe | Leu | Pro | Phe | Arg | Met | Val | Glu | Ala | Ala | Asn | Gly | Met | His |
| 65 | | | | | 70 | | | | | 75 | | | | 80 | |
| Trp | Pro | Leu | Pro | Phe | Ile | Leu | Cys | Pro | Leu | Ser | Gly | Phe | Ile | Phe | Phe |
| | | | | 85 | | | | | 90 | | | | | 95 | |

211

	Thr	Thr	Ile	Tyr	Leu	Thr	Ala	Leu	Phe	Leu	Ala	Ala	Val	Ser	Ile	Glu	
					100				105					110			
	Arg	Phe	Leu	Ser	Val	Ala	His	Pro	Leu	Trp	Tyr	Lys	Thr	Arg	Pro	Arg	
			115					120					125				
5	Leu	Gly	Gln	Ala	Gly	Leu	Val	Ser	Val	Ala	Cys	Trp	Leu	Leu	Ala	Ser	
		130					135					140					
	Ala	His	Cys	Ser	Val	Val	Tyr	Val	Ile	Glu	Phe	Ser	Gly	Asp	Ile	Ser	
	145					150					155					160	
10	His	Ser	Gln	Gly	Thr	Asn	Gly	Thr	Cys	Tyr	Leu	Glu	Phe	Arg	Lys	Asp	
					165					170					175		
	Gln	Leu	Ala	Ile	Leu	Leu	Pro	Val	Arg	Leu	Glu	Met	Ala	Val	Val	Leu	
				180					185					190			
	Phe	Val	Val	Pro	Leu	Ile	Ile	Thr	Ser	Tyr	Cys	Tyr	Ser	Arg	Leu	Val	
			195					200					205				
15	Trp	Ile	Leu	Gly	Arg	Gly	Gly	Ser	His	Arg	Arg	Gln	Arg	Arg	Val	Ala	
		210					215					220					
	Gly	Leu	Leu	Ala	Ala	Thr	Leu	Leu	Asn	Phe	Leu	Val	Cys	Phe	Gly	Pro	
	225					230					235					240	
20	Tyr	Asn	Val	Ser	His	Val	Val	Gly	Tyr	Ile	Cys	Gly	Glu	Ser	Pro	Ala	
					245					250					255		
	Trp	Arg	Ile	Tyr	Val	Thr	Leu	Leu	Ser	Thr	Leu	Asn	Ser	Cys	Val	Asp	
			260						265					270			
	Pro	Phe	Val	Tyr	Tyr	Phe	Ser	Ser	Ser	Gly	Phe	Gln	Ala	Asp	Phe	His	
			275					280					285				
25	Glu	Leu	Leu	Arg	Arg	Leu	Cys	Gly	Leu	Trp	Gly	Gln	Trp	Gln	Gln	Glu	
		290					295					300					
	Ser	Ser	Met	Glu	Leu	Lys	Glu	Gln	Lys	Gly	Gly	Glu	Glu	Gln	Arg	Ala	
	305					310					315					320	
30	Asp	Arg	Pro	Ala	Glu	Arg	Lys	Thr	Ser	Glu	His	Ser	Gln	Gly	Cys	Gly	
					325					330					335		
	Thr	Gly	Gly	Gln	Val	Ala	Cys	Ala	Glu	Ser							
					340				345								

(256) INFORMATION FOR SEQ ID NO:255:

- 35 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 31 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:255:

TTTAAGCTTC CCCTCCAGGA TGCTGCCGGA C

31

(257) INFORMATION FOR SEQ ID NO:256:

5

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: not relevant

10

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:256:

GGCGAATTCT GAAGGTCCAG GGAACTGCT A

31

(258) INFORMATION FOR SEQ ID NO:257:

15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 993 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:257:

ATGCTGCCGG ACTGGAAGAG CTCCTTGATC CTCATGGCTT ACATCATCAT CTTCTCACT 60

GGCCTCCCTG CCAACCTCCT GGCCCTGCGG GCCTTTGTGG GCGGATCCG CCAGCCCCAG 120

CCTGCACCTG TGCACATCCT CTGCTGAGC CTGACGCTGG CCGACCTCCT CTGCTGCTG 180

CTGCTGCCCT TCAAGATCAT CGAGGCTGCG TCGAACTTCC GCTGGTACCT GCCCAAGGTC 240

25 GTCTGCGCCC TCACGAGTTT TGGCTTCTAC AGCAGCATCT ACTGCAGCAC GTGGCTCCTG 300

GCGGGCATCA GCATCGAGCG CTACCTGGGA GTGGCTTTCC CCGTGCAGTA CAAGCTCTCC 360

CGCCGGCCTC TGTATGGAGT GATTGCAGCT CTGGTGGCCT GGGTTATGTC CTTTGGTCAC 420

TGCACCATCG TGATCATCGT TCAATACTTG AACACGACTG AGCAGGTCAG AAGTGGCAAT 480

GAAATTACCT GCTACGAGAA CTTACCGAT AACCAGTTGG ACGTGGTGCT GCCCGTCCG 540

30 CTGGAGCTGT GCCTGGTGCT CTTCTTCATC CCCATGGCAG TCACCATCTT CTGCTACTGG 600

CGTTTGTGT GGATCATGCT CTCCCAGCCC CTTGTGGGGG CCCAGAGGCG GCGCCGAGCC 660

GTGGGGCTGG CTGTGGTGAC GCTGCTCAAT TTCCTGGTGT GCTTCGGACC TTACAACGTG 720

213

TCCCACCTGG TGGGGTATCA CCAGAGAAAA AGCCCCTGGT GCGGTCAAT AGCCGTGGTG 780
 TTCAGTTCAC TCAACGCCAG TCTGGACCCC CTGCTCTTCT ATTTCTCTTC TTCAGTGGTG 840
 CGCAGGGCAT TTGGGAGAGG GCTGCAGGTG CTGCGGAATC AGGGCTCCTC CCTGTTGGGA 900
 CGCAGAGGCA AAGACACAGC AGAGGGGACA AATGAGGACA GGGGTGTGGG TCAAGGAGAA 960
 5 GGGATGCCAA GTTCGGACTT CACTACAGAG TAG 993

(259) INFORMATION FOR SEQ ID NO:258:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 362 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:258:

15 Met Leu Pro Asp Trp Lys Ser Ser Leu Ile Leu Met Ala Tyr Ile Ile
 1 5 10 15
 Ile Phe Leu Thr Gly Leu Pro Ala Asn Leu Leu Ala Leu Arg Ala Phe
 20 25 30
 Val Gly Arg Ile Arg Gln Pro Gln Pro Ala Pro Val His Ile Leu Leu
 35 40 45
 20 Leu Ser Leu Thr Leu Ala Asp Leu Leu Leu Leu Leu Leu Leu Pro Phe
 50 55 60
 Lys Ile Ile Glu Ala Ala Ser Asn Phe Arg Trp Tyr Leu Pro Lys Val
 65 70 75 80
 25 Val Cys Ala Leu Thr Ser Phe Gly Phe Tyr Ser Ser Ile Tyr Cys Ser
 85 90 95
 Thr Trp Leu Leu Ala Gly Ile Ser Ile Glu Arg Tyr Leu Gly Val Ala
 100 105 110
 Phe Pro Val Gln Tyr Lys Leu Ser Arg Arg Pro Leu Tyr Gly Val Ile
 115 120 125
 30 Ala Ala Leu Val Ala Trp Val Met Ser Phe Gly His Cys Thr Ile Val
 130 135 140
 Ile Ile Val Gln Tyr Leu Asn Thr Thr Glu Gln Val Arg Ser Gly Asn
 145 150 155 160
 35 Glu Ile Thr Cys Tyr Glu Asn Phe Thr Asp Asn Gln Leu Asp Val Val
 165 170 175

Leu Pro Val Arg Leu Glu Leu Cys Leu Val Leu Phe Phe Ile Pro Met
 180 185 190
 Ala Val Thr Ile Phe Cys Tyr Trp Arg Phe Val Trp Ile Met Leu Ser
 195 200 205
 5 Gln Pro Leu Val Gly Ala Gln Arg Arg Arg Arg Ala Val Gly Leu Ala
 210 215 220
 Val Val Thr Leu Leu Asn Phe Leu Val Cys Phe Gly Pro Tyr Asn Val
 225 230 235 240
 10 Ser His Leu Val Gly Tyr His Gln Arg Lys Ser Pro Trp Trp Arg Ser
 245 250 255
 Ile Ala Val Val Phe Ser Ser Leu Asn Ala Ser Leu Asp Pro Leu Leu
 260 265 270
 Phe Tyr Phe Ser Ser Ser Val Val Arg Arg Ala Phe Gly Arg Gly Leu
 275 280 285
 15 Gln Val Leu Arg Asn Gln Gly Ser Ser Leu Leu Gly Arg Arg Gly Lys
 290 295 300
 Asp Thr Ala Glu Gly Thr Asn Glu Asp Arg Gly Val Gly Gln Gly Glu
 305 310 315 320
 20 Gly Met Pro Ser Ser Asp Phe Thr Thr Glu
 325 330

(260) INFORMATION FOR SEQ ID NO:259:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 25 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: DNA (genomic)
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:259:

CCCAAGCTTC GGGCACCATG GACACCTCCC

30

30 (261) INFORMATION FOR SEQ ID NO:260:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 35 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: DNA (genomic)
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:260:

ACAGGATCCA AATGCACAGC ACTGGTAAGC

30

(262) INFORMATION FOR SEQ ID NO:261:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:261:

10 CTATAACTGG GTTACATGGT TTAAC

25

(263) INFORMATION FOR SEQ ID NO:262:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:262:

TTTGAATTCA CATATTAATT AGAGACATGG

30

20 (264) INFORMATION FOR SEQ ID NO:263:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2724 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:263:

ATGGACACCT CCCGGCTCGG TGTGCTCCTG TCCTTGCCTG TGCTGCTGCA GCTGGCGACC 60

GGGGGCAGCT CTCCCAGGTC TGGTGTGTTG CTGAGGGGCT GCCCCACACA CTGTCATTGC 120

30 GAGCCCGACG GCAGGATGTT GCTCAGGGTG GACTGCTCCG ACCTGGGGCT CTCGGAGCTG 180

CCTTCCAACC TCAGCGTCTT CACCTCCTAC CTAGACCTCA GTATGAACAA CATCAGTCAG 240

CTGCTCCCGA ATCCCCTGCC CAGTCTCCGC TTCCTGGAGG AGTTACGTCT TGCGGGAAAC 300

GCTCTGACAT ACATTCCCAA GGGAGCATTC ACTGGCCTTT ACAGTCTTAA AGTTCTTATG 360

CTGCAGAATA ATCAGCTAAG ACACGTACCC ACAGAAGCTC TGCAGAATTT GCGAAGCCTT 420
 CAATCCCTGC GTCTGGATGC TAACCACATC AGCTATGTGC CCCCAAGCTG TTTCAGTGGC 480
 CTGCATTCCC TGAGGCACCT GTGGCTGGAT GACAATGCGT TAACAGAAAT CCCCCTCCAG 540
 GCTTTTAGAA GTTTATCGGC ATTGCAAGCC ATGACCTTGG CCCTGAACAA AATACACCAC 600
 5 ATACCAGACT ATGCCTTTGG AAACCTCTCC AGCTTGGTAG TTCTACATCT CCATAACAAT 660
 AGAATCCACT CCCTGGGAAA GAAATGCTTT GATGGGCTCC ACAGCCTAGA GACTTTAGAT 720
 TTAAATTACA ATAACCTTGA TGAATTCCCC ACTGCAATTA GGACACTCTC CAACCTTAAA 780
 GAACTAGGAT TTCATAGCAA CAATATCAGG TCGATACCTG AGAAAGCATT TGTAGGCAAC 840
 CTTCTCTTA TTACAATACA TTTCTATGAC AATCCCATCC AATTTGTTGG GAGATCTGCT 900
 10 TTTCACATT TACCTGAAGT AAGAACACTG ACTCTGAATG GTGCCTCACA AATAACTGAA 960
 TTTCCTGATT TAACTGGAAC TGCAAACCTG GAGAGTCTGA CTTTAACTGG AGCACAGATC 1020
 TCATCTCTTC CTCAAACCGT CTGCAATCAG TTACCTAATC TCCAAGTGCT AGATCTGTCT 1080
 TACAACCTAT TAGAAGATTT ACCCAGTTTT TCAGTCTGCC AAAAGCTTCA GAAAATTGAC 1140
 CTAAGACATA ATGAAATCTA CGAAATTAAA GTTGACACTT TCCAGCAGTT GCTTAGCCTC 1200
 15 CGATCGCTGA ATTTGGCTTG GAACAAAATT GCTATTATTC ACCCCAATGC ATTTTCCACT 1260
 TTGCCATCCC TAATAAGCT GGACCTATCG TCCAACCTCC TGTCGTCTTT TCCTATAACT 1320
 GGGTTACATG GTTTAACTCA CTTAAAATTA ACAGGAAATC ATGCCTTACA GAGCTTGATA 1380
 TCATCTGAAA ACTTTCCAGA ACTCAAGGTT ATAGAAATGC CTTATGCTTA CCAGTGCTGT 1440
 GCATTTGGAG TGTGTGAGAA TGCCTATAAG ATTTCTAATC AATGGAATAA AGGTGACAAC 1500
 20 AGCAGTATGG ACGACCTTCA TAAGAAAGAT GCTGGAATGT TTCAGGCTCA AGATGAACGT 1560
 GACCTTGAAG ATTTCTGCT TGACTTTGAG GAAGACCTGA AAGCCCTTCA TTCAGTGCAG 1620
 TGTTACCTT CCCCAGGCCC CTTCAAACCC TGTGAACACC TGCTTGATGG CTGGCTGATC 1680
 AGAATTGGAG TGTGGACCAT AGCAGTTCTG GCACTTACTT GTAATGCTTT GGTGACTTCA 1740
 ACAGTTTTCA GATCCCCTCT GTACATTTCC CCCATTAAAC TGTTAATTGG GGTCATCGCA 1800
 25 GCAGTGAACA TGCTCACGGG AGTCTCCAGT GCCGTGCTGG CTGGTGTGGA TCGGTTCACT 1860
 TTTGGCAGCT TTGCACGACA TGGTGCCTGG TGGGAGAATG GGGTTGGTTG CCATGTCTAT 1920
 GGTTTTTTGT CCATTTTTGC TTCAGAATCA TCTGTTTTCC TGCTTACTCT GGCAGCCCTG 1980
 GAGCGTGGGT TCTCTGTGAA ATATTCTGCA AAATTTGAAA CGAAAGCTCC ATTTTCTAGC 2040

217

CTGAAAGTAA TCATTTTGCT CTGTGCCCTG CTGGCCTTGA CCATGGCCGC AGTTCCTCTG 2100
 CTGGGTGGCA GCAAGTATGG CGCTCCCCT CTCTGCCTGC CTTTGCCTTT TGGGGAGCCC 2160
 AGCACCATGG GCTACATGGT CGCTCTCATC TTGCTCAATT CCCTTTGCTT CCTCATGATG 2220
 ACCATTGCCT ACACCAAGCT CTA CTGCAAT TTGGACAAGG GAGACCTGGA GAATATTTGG 2280
 5 GACTGCTCTA TGGTAAAACA CATTGCCCTG TTGCTCTTCA CCAACTGCAT CCTAAACTGC 2340
 CCTGTGGCTT TCTTGTCTT CTCCTCTTTA ATAAACCTTA CATTTATCAG TCCTGAAGTA 2400
 ATTAAGTTTA TCCTTCTGGT GGTAGTCCCA CTTCCTGCAT GTCTCAATCC CCTTCTCTAC 2460
 ATCTTGTTCA ATCCTCACTT TAAGGAGGAT CTGGTGAGCC TGAGAAAGCA AACCTACGTC 2520
 TGGACAAGAT CAAAACACCC AAGCTTGATG TCAATTAAC CTGATGATGT CGAAAAACAG 2580
 10 TCCTGTGACT CAACTCAAGC CTTGGTAACC TTTACCAGCT CCAGCATCAC TTATGACCTG 2640
 CCTCCCAGTT CCGTGCCATC ACCAGCTTAT CCAGTGACTG AGAGCTGCCA TCTTTCCTCT 2700
 GTGGCATTG TCCCATGTCT CTAA 2724

(265) INFORMATION FOR SEQ ID NO:264:

- 15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 907 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

- 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:264:

Met Asp Thr Ser Arg Leu Gly Val Leu Leu Ser Leu Pro Val Leu Leu
 1 5 10 15

Gln Leu Ala Thr Gly Gly Ser Ser Pro Arg Ser Gly Val Leu Leu Arg
 20 25 30

25 Gly Cys Pro Thr His Cys His Cys Glu Pro Asp Gly Arg Met Leu Leu
 35 40 45

Arg Val Asp Cys Ser Asp Leu Gly Leu Ser Glu Leu Pro Ser Asn Leu
 50 55 60

30 Ser Val Phe Thr Ser Tyr Leu Asp Leu Ser Met Asn Asn Ile Ser Gln
 65 70 75 80

Leu Leu Pro Asn Pro Leu Pro Ser Leu Arg Phe Leu Glu Glu Leu Arg
 85 90 95

Leu Ala Gly Asn Ala Leu Thr Tyr Ile Pro Lys Gly Ala Phe Thr Gly

	100		105		110
	Leu Tyr Ser Leu Lys Val Leu Met Leu Gln Asn Asn Gln Leu Arg His	115	120		125
5	Val Pro Thr Glu Ala Leu Gln Asn Leu Arg Ser Leu Gln Ser Leu Arg	130	135	140	
	Leu Asp Ala Asn His Ile Ser Tyr Val Pro Pro Ser Cys Phe Ser Gly	145	150	155	160
	Leu His Ser Leu Arg His Leu Trp Leu Asp Asp Asn Ala Leu Thr Glu	165	170		175
10	Ile Pro Val Gln Ala Phe Arg Ser Leu Ser Ala Leu Gln Ala Met Thr	180	185		190
	Leu Ala Leu Asn Lys Ile His His Ile Pro Asp Tyr Ala Phe Gly Asn	195	200		205
15	Leu Ser Ser Leu Val Val Leu His Leu His Asn Asn Arg Ile His Ser	210	215	220	
	Leu Gly Lys Lys Cys Phe Asp Gly Leu His Ser Leu Glu Thr Leu Asp	225	230	235	240
	Leu Asn Tyr Asn Asn Leu Asp Glu Phe Pro Thr Ala Ile Arg Thr Leu	245	250		255
20	Ser Asn Leu Lys Glu Leu Gly Phe His Ser Asn Asn Ile Arg Ser Ile	260	265		270
	Pro Glu Lys Ala Phe Val Gly Asn Pro Ser Leu Ile Thr Ile His Phe	275	280		285
25	Tyr Asp Asn Pro Ile Gln Phe Val Gly Arg Ser Ala Phe Gln His Leu	290	295	300	
	Pro Glu Leu Arg Thr Leu Thr Leu Asn Gly Ala Ser Gln Ile Thr Glu	305	310	315	320
	Phe Pro Asp Leu Thr Gly Thr Ala Asn Leu Glu Ser Leu Thr Leu Thr	325	330		335
30	Gly Ala Gln Ile Ser Ser Leu Pro Gln Thr Val Cys Asn Gln Leu Pro	340	345		350
	Asn Leu Gln Val Leu Asp Leu Ser Tyr Asn Leu Leu Glu Asp Leu Pro	355	360		365
35	Ser Phe Ser Val Cys Gln Lys Leu Gln Lys Ile Asp Leu Arg His Asn	370	375	380	
	Glu Ile Tyr Glu Ile Lys Val Asp Thr Phe Gln Gln Leu Leu Ser Leu	385	390	395	400

219

	Arg	Ser	Leu	Asn	Leu	Ala	Trp	Asn	Lys	Ile	Ala	Ile	Ile	His	Pro	Asn	
					405					410					415		
	Ala	Phe	Ser	Thr	Leu	Pro	Ser	Leu	Ile	Lys	Leu	Asp	Leu	Ser	Ser	Asn	
				420					425					430			
5	Leu	Leu	Ser	Ser	Phe	Pro	Ile	Thr	Gly	Leu	His	Gly	Leu	Thr	His	Leu	
			435					440					445				
	Lys	Leu	Thr	Gly	Asn	His	Ala	Leu	Gln	Ser	Leu	Ile	Ser	Ser	Glu	Asn	
		450					455					460					
10	Phe	Pro	Glu	Leu	Lys	Val	Ile	Glu	Met	Pro	Tyr	Ala	Tyr	Gln	Cys	Cys	
	465					470					475					480	
	Ala	Phe	Gly	Val	Cys	Glu	Asn	Ala	Tyr	Lys	Ile	Ser	Asn	Gln	Trp	Asn	
				485						490					495		
	Lys	Gly	Asp	Asn	Ser	Ser	Met	Asp	Asp	Leu	His	Lys	Lys	Asp	Ala	Gly	
				500					505					510			
15	Met	Phe	Gln	Ala	Gln	Asp	Glu	Arg	Asp	Leu	Glu	Asp	Phe	Leu	Leu	Asp	
			515					520					525				
	Phe	Glu	Glu	Asp	Leu	Lys	Ala	Leu	His	Ser	Val	Gln	Cys	Ser	Pro	Ser	
		530					535					540					
20	Pro	Gly	Pro	Phe	Lys	Pro	Cys	Glu	His	Leu	Leu	Asp	Gly	Trp	Leu	Ile	
	545					550					555					560	
	Arg	Ile	Gly	Val	Trp	Thr	Ile	Ala	Val	Leu	Ala	Leu	Thr	Cys	Asn	Ala	
				565						570					575		
	Leu	Val	Thr	Ser	Thr	Val	Phe	Arg	Ser	Pro	Leu	Tyr	Ile	Ser	Pro	Ile	
				580					585					590			
25	Lys	Leu	Leu	Ile	Gly	Val	Ile	Ala	Ala	Val	Asn	Met	Leu	Thr	Gly	Val	
		595						600					605				
	Ser	Ser	Ala	Val	Leu	Ala	Gly	Val	Asp	Ala	Phe	Thr	Phe	Gly	Ser	Phe	
		610					615					620					
30	Ala	Arg	His	Gly	Ala	Trp	Trp	Glu	Asn	Gly	Val	Gly	Cys	His	Val	Ile	
	625					630					635					640	
	Gly	Phe	Leu	Ser	Ile	Phe	Ala	Ser	Glu	Ser	Ser	Val	Phe	Leu	Leu	Thr	
				645						650				655			
	Leu	Ala	Ala	Leu	Glu	Arg	Gly	Phe	Ser	Val	Lys	Tyr	Ser	Ala	Lys	Phe	
				660					665					670			
35	Glu	Thr	Lys	Ala	Pro	Phe	Ser	Ser	Leu	Lys	Val	Ile	Ile	Leu	Leu	Cys	
		675						680					685				
	Ala	Leu	Leu	Ala	Leu	Thr	Met	Ala	Ala	Val	Pro	Leu	Leu	Gly	Gly	Ser	

	690		695		700
	Lys Tyr Gly Ala Ser Pro Leu Cys Leu Pro Leu Pro Phe Gly Glu Pro				
	705		710		715 720
5	Ser Thr Met Gly Tyr Met Val Ala Leu Ile Leu Leu Asn Ser Leu Cys				
		725		730	735
	Phe Leu Met Met Thr Ile Ala Tyr Thr Lys Leu Tyr Cys Asn Leu Asp				
		740		745	750
	Lys Gly Asp Leu Glu Asn Ile Trp Asp Cys Ser Met Val Lys His Ile				
		755		760	765
10	Ala Leu Leu Leu Phe Thr Asn Cys Ile Leu Asn Cys Pro Val Ala Phe				
		770		775	780
	Leu Ser Phe Ser Ser Leu Ile Asn Leu Thr Phe Ile Ser Pro Glu Val				
		785		790 795	800
15	Ile Lys Phe Ile Leu Leu Val Val Val Pro Leu Pro Ala Cys Leu Asn				
		805		810	815
	Pro Leu Leu Tyr Ile Leu Phe Asn Pro His Phe Lys Glu Asp Leu Val				
		820		825	830
	Ser Leu Arg Lys Gln Thr Tyr Val Trp Thr Arg Ser Lys His Pro Ser				
		835		840	845
20	Leu Met Ser Ile Asn Ser Asp Asp Val Glu Lys Gln Ser Cys Asp Ser				
		850		855	860
	Thr Gln Ala Leu Val Thr Phe Thr Ser Ser Ser Ile Thr Tyr Asp Leu				
		865		870 875	880
25	Pro Pro Ser Ser Val Pro Ser Pro Ala Tyr Pro Val Thr Glu Ser Cys				
		885		890	895
	His Leu Ser Ser Val Ala Phe Val Pro Cys Leu				
		900		905	

(266) INFORMATION FOR SEQ ID NO:265:

- 30 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- 35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:265:

CGGAAGCTGC GGGCCAAATG GGTGGCCGGC

221

(267) INFORMATION FOR SEQ ID NO:266:

- 5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:266:

CAGAGGAGGG TGAAGGGGCT GTTGGCG

27

10 (268) INFORMATION FOR SEQ ID NO:267:

- 15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:267:

GGCGGCGCCG AGCCAAGGGG CTGGCTGTGG

30

(269) INFORMATION FOR SEQ ID NO:268:

- 20 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- 25 (ii) MOLECULE TYPE: DNA (genomic)
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:268:

GGGACTGCTC TATGAAAAAA CACATTGCCC TG

32

(270) INFORMATION FOR SEQ ID NO:269:

- 30 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1071 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:269:

ATGAATGGGG TCTCGGAGGG GACCAGAGGC TGCAGTGACA GGCAACCTGG GGTCTGACA

60

CGTGATCGCT CTTGTTCCAG GAAGATGAAC TCTTCCGGAT GCCTGTCTGA GGAGGTGGGG 120
TCCCTCCGCC CACTGACTGT GGTATCCTG TCTGCGTCCA TTGTCGTCGG AGTGCTGGGC 180
AATGGGCTGG TGCTGTGGAT GACTGTCTTC CGTATGGCAC GCACGGTCTC CACCGTCTGC 240
TTCTTCCACC TGGCCCTTGC CGATTTCATG CTCTCACTGT CTCTGCCCAT TGCCATGTAC 300
5 TATATTGTCT CCAGGCAGTG GCTCCTCGGA GAGTGGGCCT GCAAACCTCTA CATCACCTTT 360
GTGTTCTCTCA GCTACTTTGC CAGTAACTGC CTCCTTGTCT TCATCTCTGT GGACCGTTGC 420
ATCTCTGTCC TCTACCCCGT CTGGGCCCTG AACCACCGCA CTGTGCAGCG GGCGAGCTGG 480
CTGGCCTTTG GGGTGTGGCT CCTGGCCGCC GCCTTGTGCT CTGCGCACCT GAAATTCCGG 540
ACAACCAGAA AATGGAATGG CTGTACGCAC TGCTACTTGG CGTTCAACTC TGACAATGAG 600
10 ACTGCCCAGA TTTGGATTGA AGGGGTCGTG GAGGGACACA TTATAGGGAC CATTGGCCAC 660
TTCCTGCTGG GCTTCCTGGG GCCCTTAGCA ATCATAGGCA CCTGCGCCCA CCTCATCCGG 720
GCCAAGCTCT TCGGGGAGGG CTGGGTCCAT GCCAACCAGC CCAAGAGGCT GCTGCTGGTG 780
CTGGTGAGCG CTTTCTTTAT CTTCTGGTCC CCGTTTAACG TGGTGCTGTT GGTCCATCTG 840
TGCGGACGGG TGATGCTCAA GGAAATCTAC CACCCCGGA TGCTGCTCAT CCTCCAGGCT 900
15 AGCTTTGCCT TGGGCTGTGT CAACAGCAGC CTCAACCCCT TCCTCTACGT CTTCGTTGGC 960
AGAGATTTC AAGAAAAGTT TTTCCAGTCT TTGACTTCTG CCCTGGCGAG GCGGTTTGGA 1020
GAGGAGGAGT TTCTGTCATC CTGTCCCCGT GGCAACGCCC CCCGGGAATG A 1071

(271) INFORMATION FOR SEQ ID NO:270:

- 20 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 356 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

- 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:270:

Met Asn Gly Val Ser Glu Gly Thr Arg Gly Cys Ser Asp Arg Gln Pro
1 5 10 15

Gly Val Leu Thr Arg Asp Arg Ser Cys Ser Arg Lys Met Asn Ser Ser
20 25 30

30 Gly Cys Leu Ser Glu Glu Val Gly Ser Leu Arg Pro Leu Thr Val Val
35 40 45

223

	Ile	Leu	Ser	Ala	Ser	Ile	Val	Val	Gly	Val	Leu	Gly	Asn	Gly	Leu	Val	
	50					55					60						
	Leu	Trp	Met	Thr	Val	Phe	Arg	Met	Ala	Arg	Thr	Val	Ser	Thr	Val	Cys	
	65					70					75					80	
5	Phe	Phe	His	Leu	Ala	Leu	Ala	Asp	Phe	Met	Leu	Ser	Leu	Ser	Leu	Pro	
				85					90						95		
	Ile	Ala	Met	Tyr	Tyr	Ile	Val	Ser	Arg	Gln	Trp	Leu	Leu	Gly	Glu	Trp	
				100					105						110		
10	Ala	Cys	Lys	Leu	Tyr	Ile	Thr	Phe	Val	Phe	Leu	Ser	Tyr	Phe	Ala	Ser	
			115					120					125				
	Asn	Cys	Leu	Leu	Val	Phe	Ile	Ser	Val	Asp	Arg	Cys	Ile	Ser	Val	Leu	
		130				135						140					
	Tyr	Pro	Val	Trp	Ala	Leu	Asn	His	Arg	Thr	Val	Gln	Arg	Ala	Ser	Trp	
	145					150					155					160	
15	Leu	Ala	Phe	Gly	Val	Trp	Leu	Leu	Ala	Ala	Ala	Leu	Cys	Ser	Ala	His	
				165					170						175		
	Leu	Lys	Phe	Arg	Thr	Thr	Arg	Lys	Trp	Asn	Gly	Cys	Thr	His	Cys	Tyr	
				180					185					190			
20	Leu	Ala	Phe	Asn	Ser	Asp	Asn	Glu	Thr	Ala	Gln	Ile	Trp	Ile	Glu	Gly	
			195			200							205				
	Val	Val	Glu	Gly	His	Ile	Ile	Gly	Thr	Ile	Gly	His	Phe	Leu	Leu	Gly	
		210				215						220					
	Phe	Leu	Gly	Pro	Leu	Ala	Ile	Ile	Gly	Thr	Cys	Ala	His	Leu	Ile	Arg	
	225					230					235					240	
25	Ala	Lys	Leu	Leu	Arg	Glu	Gly	Trp	Val	His	Ala	Asn	Arg	Pro	Lys	Arg	
				245					250					255			
	Leu	Leu	Leu	Val	Leu	Val	Ser	Ala	Phe	Phe	Ile	Phe	Trp	Ser	Pro	Phe	
				260					265					270			
30	Asn	Val	Val	Leu	Leu	Val	His	Leu	Trp	Arg	Arg	Val	Met	Leu	Lys	Glu	
		275				280							285				
	Ile	Tyr	His	Pro	Arg	Met	Leu	Leu	Ile	Leu	Gln	Ala	Ser	Phe	Ala	Leu	
		290				295						300					
	Gly	Cys	Val	Asn	Ser	Ser	Leu	Asn	Pro	Phe	Leu	Tyr	Val	Phe	Val	Gly	
	305					310					315					320	
35	Arg	Asp	Phe	Gln	Glu	Lys	Phe	Phe	Gln	Ser	Leu	Thr	Ser	Ala	Leu	Ala	
				325					330						335		
	Arg	Ala	Phe	Gly	Glu	Glu	Glu	Phe	Leu	Ser	Ser	Cys	Pro	Arg	Gly	Asn	

340

345

350

Ala Pro Arg Glu
355

(272) INFORMATION FOR SEQ ID NO:271:

5

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 903 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:271:

```

ATGGACCTGC CCCCAGCT CTCCTTCGGC CTCTATGTGG CCGCCTTTGC GCTGGGCTTC      60
CCGCTCAACG TCCTGGCCAT CCGAGGCGCG ACGGCCACG CCCGGCTCCG TCTACCCCT      120
15 AGCCTGGTCT ACGCCCTGAA CCTGGGCTGC TCCGACCTGC TGCTGACAGT CTCTCTGCCC      180
CTGAAGGCGG TGGAGGCGCT AGCCTCCGGG GCCTGGCCTC TGCCGGCCTC GCTGTGCCCC      240
GTCTTCGCGG TGGCCCACTT CTTCCCACTC TATGCCGCGG GGGGCTTCCT GGCCGCCCTG      300
AGTGCAGGCC GCTACCTGGG AGCAGCCTTC CCCTTGGGCT ACCAAGCCTT CCGGAGGCCG      360
TGCTATTCTT GGGGGGTGTG CGCGGCCATC TGGGCCCTCG TCCTGTGTCA CCTGGGTCTG      420
20 GTCTTTGGGT TGGAGGCTCC AGGAGGCTGG CTGGACCACA GCAACACCTC CCTGGGCATC      480
AACACACCGG TCAACGGCTC TCCGGTCTGC CTGGAGGCCT GGGACCCGGC CTCTGCCGGC      540
CCGGCCCGCT TCAGCCTCTC TCTCCTGCTC TTTTCTGCTG CCTTGGCCAT CACAGCCTTC      600
TGCTACGTGG GCTGCCTCCG GGCCTGGCC CGCTCCGGCC TGACGCACAG GCGGAAGCTG      660
CGGGCCAAAT GGGTGGCCGG CGGGGCCCTC CTCACGCTGC TGCTCTGCGT AGGACCCTAC      720
25 AACGCCTCCA ACGTGGCCAG CTTCTGTAC CCCAATCTAG GAGGCTCCTG GCGGAAGCTG      780
GGGCTCATCA CGGGTGCCTG GAGTGTGGTG CTTAATCCGC TGGTGACCGG TTAATTGGGA      840
AGGGGTCTTG GCCTGAAGAC AGTGTGTGCG GCAAGAACGC AAGGGGGCAA GTCCCAGAAG      900
TAA                                                                    903

```

(273) INFORMATION FOR SEQ ID NO:272:

30

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 300 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

225

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:272:

5	Met	Asp	Leu	Pro	Pro	Gln	Leu	Ser	Phe	Gly	Leu	Tyr	Val	Ala	Ala	Phe	1	5	10	15
	Ala	Leu	Gly	Phe	Pro	Leu	Asn	Val	Leu	Ala	Ile	Arg	Gly	Ala	Thr	Ala	20	25	30	
	His	Ala	Arg	Leu	Arg	Leu	Thr	Pro	Ser	Leu	Val	Tyr	Ala	Leu	Asn	Leu	35	40	45	
10	Gly	Cys	Ser	Asp	Leu	Leu	Leu	Thr	Val	Ser	Leu	Pro	Leu	Lys	Ala	Val	50	55	60	
	Glu	Ala	Leu	Ala	Ser	Gly	Ala	Trp	Pro	Leu	Pro	Ala	Ser	Leu	Cys	Pro	65	70	75	80
15	Val	Phe	Ala	Val	Ala	His	Phe	Phe	Pro	Leu	Tyr	Ala	Gly	Gly	Gly	Phe	85	90	95	
	Leu	Ala	Ala	Leu	Ser	Ala	Gly	Arg	Tyr	Leu	Gly	Ala	Ala	Phe	Pro	Leu	100	105	110	
	Gly	Tyr	Gln	Ala	Phe	Arg	Arg	Pro	Cys	Tyr	Ser	Trp	Gly	Val	Cys	Ala	115	120	125	
20	Ala	Ile	Trp	Ala	Leu	Val	Leu	Cys	His	Leu	Gly	Leu	Val	Phe	Gly	Leu	130	135	140	
	Glu	Ala	Pro	Gly	Gly	Trp	Leu	Asp	His	Ser	Asn	Thr	Ser	Leu	Gly	Ile	145	150	155	160
25	Asn	Thr	Pro	Val	Asn	Gly	Ser	Pro	Val	Cys	Leu	Glu	Ala	Trp	Asp	Pro	165	170	175	
	Ala	Ser	Ala	Gly	Pro	Ala	Arg	Phe	Ser	Leu	Ser	Leu	Leu	Leu	Phe	Phe	180	185	190	
	Leu	Pro	Leu	Ala	Ile	Thr	Ala	Phe	Cys	Tyr	Val	Gly	Cys	Leu	Arg	Ala	195	200	205	
30	Leu	Ala	Arg	Ser	Gly	Leu	Thr	His	Arg	Arg	Lys	Leu	Arg	Ala	Lys	Trp	210	215	220	
	Val	Ala	Gly	Gly	Ala	Leu	Leu	Thr	Leu	Leu	Leu	Cys	Val	Gly	Pro	Tyr	225	230	235	240
35	Asn	Ala	Ser	Asn	Val	Ala	Ser	Phe	Leu	Tyr	Pro	Asn	Leu	Gly	Gly	Ser	245	250	255	
	Trp	Arg	Lys	Leu	Gly	Leu	Ile	Thr	Gly	Ala	Trp	Ser	Val	Val	Leu	Asn				

260

265

270

Pro Leu Val Thr Gly Tyr Leu Gly Arg Gly Pro Gly Leu Lys Thr Val
 275 280 285

5 Cys Ala Ala Arg Thr Gln Gly Gly Lys Ser Gln Lys
 290 295 300

(274). INFORMATION FOR SEQ ID NO:273:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1041 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:273:

10 ATGGATACAG GCCCGACCA GTCCTACTTC TCCGGCAATC ACTGGTTCGT CTTCTCGGTG 60

15 TACCTTCTCA CTTTCTGGT GGGGCTCCCC CTCAACCTGC TGGCCCTGGT GGTCTTCGTG 120

GGCAAGCTGC AGCGCCGCCC GGTGGCCGTG GACGTGCTCC TGCTCAACCT GACCGCCTCG 180

GACCTGCTCC TGCTGCTGTT CCTGCCTTTC CGCATGGTGG AGGCAGCCAA TGGCATGCAC 240

TGGCCCTGC CCTTCATCCT CTGCCCACTC TCTGGATTCA TCTTCTTCAC CACCATCTAT 300

CTCACCGCCC TCTTCTGGC AGCTGTGAGC ATTGAACGCT TCCTGAGTGT GGCCACCCA 360

20 CTGTGGTACA AGACCCGGCC GAGGCTGGGG CAGGCAGGTC TGGTGAGTGT GGCCTGCTGG 420

CTGTTGGCCT CTGCTCACTG CAGCGTGGTC TACGTCATAG AATTCTCAGG GGACATCTCC 480

CACAGCCAGG GCACCAATGG GACCTGCTAC CTGGAGTTCC GGAAGGACCA GCTAGCCATC 540

CTCCTGCCCC TGCGGCTGGA GATGGCTGTG GTCCTCTTTG TGGTCCCGCT GATCATCACC 600

AGCTACTGCT ACAGCCGCCT GGTGTGGATC CTCGGCAGAG GGGGCAGCCA CCGCCGGCAG 660

25 AGGAGGGTGA AGGGGCTGTT GGCGGCCACG CTGCTCAACT TCCTTGTCTG CTTTGGGCCC 720

TACAACGTGT CCCATGTCGT GGGCTATATC TGCGGTGAAA GCGCGCATG GAGGATCTAC 780

GTGACGCTTC TCAGCACCCT GAACTCCTGT GTCGACCCCT TTGTCTACTA CTTCTCCTCC 840

TCCGGGTTCC AAGCCGACTT TCATGAGCTG CTGAGGAGGT TGTGTGGGCT CTGGGGCCAG 900

TGGCAGCAGG AGAGCAGCAT GGAGCTGAAG GAGCAGAAGG GAGGGGAGGA GCAGAGAGCG 960

30 GACCGACCAG CTGAAAGAAA GACCAAGTAA CACTCACAGG GCTGTGGAAC TGGTGGCCAG 1020

GTGGCCTGTG CTGAAAGCTA G 1041

227

(275) INFORMATION FOR SEQ ID NO:274:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 346 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:274:

5 Met Asp Thr Gly Pro Asp Gln Ser Tyr Phe Ser Gly Asn His Trp Phe
 1 5 10 15
 Val Phe Ser Val Tyr Leu Leu Thr Phe Leu Val Gly Leu Pro Leu Asn
 20 25 30
 Leu Leu Ala Leu Val Val Phe Val Gly Lys Leu Gln Arg Arg Pro Val
 35 40 45
 15 Ala Val Asp Val Leu Leu Leu Asn Leu Thr Ala Ser Asp Leu Leu Leu
 50 55 60
 Leu Leu Phe Leu Pro Phe Arg Met Val Glu Ala Ala Asn Gly Met His
 65 70 75 80
 20 Trp Pro Leu Pro Phe Ile Leu Cys Pro Leu Ser Gly Phe Ile Phe Phe
 85 90 95
 Thr Thr Ile Tyr Leu Thr Ala Leu Phe Leu Ala Ala Val Ser Ile Glu
 100 105 110
 Arg Phe Leu Ser Val Ala His Pro Leu Trp Tyr Lys Thr Arg Pro Arg
 115 120 125
 25 Leu Gly Gln Ala Gly Leu Val Ser Val Ala Cys Trp Leu Leu Ala Ser
 130 135 140
 Ala His Cys Ser Val Val Tyr Val Ile Glu Phe Ser Gly Asp Ile Ser
 145 150 155 160
 30 His Ser Gln Gly Thr Asn Gly Thr Cys Tyr Leu Glu Phe Arg Lys Asp
 165 170 175
 Gln Leu Ala Ile Leu Leu Pro Val Arg Leu Glu Met Ala Val Val Leu
 180 185 190
 Phe Val Val Pro Leu Ile Ile Thr Ser Tyr Cys Tyr Ser Arg Leu Val
 195 200 205
 35 Trp Ile Leu Gly Arg Gly Gly Ser His Arg Arg Gln Arg Arg Val Lys
 210 215 220
 Gly Leu Leu Ala Ala Thr Leu Leu Asn Phe Leu Val Cys Phe Gly Pro

	225		230		235		240									
	Tyr	Asn	Val	Ser	His	Val	Val	Gly	Tyr	Ile	Cys	Gly	Glu	Ser	Pro	Ala
					245					250					255	
5	Trp	Arg	Ile	Tyr	Val	Thr	Leu	Leu	Ser	Thr	Leu	Asn	Ser	Cys	Val	Asp
				260					265					270		
	Pro	Phe	Val	Tyr	Tyr	Phe	Ser	Ser	Ser	Gly	Phe	Gln	Ala	Asp	Phe	His
			275						280				285			
	Glu	Leu	Leu	Arg	Arg	Leu	Cys	Gly	Leu	Trp	Gly	Gln	Trp	Gln	Gln	Glu
		290					295					300				
10	Ser	Ser	Met	Glu	Leu	Lys	Glu	Gln	Lys	Gly	Gly	Glu	Glu	Gln	Arg	Ala
	305					310					315				320	
	Asp	Arg	Pro	Ala	Glu	Arg	Lys	Thr	Ser	Glu	His	Ser	Gln	Gly	Cys	Gly
				325					330					335		
15	Thr	Gly	Gly	Gln	Val	Ala	Cys	Ala	Glu	Ser						
			340						345							

(276) INFORMATION FOR SEQ ID NO:275:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 993 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:275:

	ATGCTGCCGG	ACTGGAAGAG	CTCCTTGATC	CTCATGGCTT	ACATCATCAT	CTTCCTCACT	60
25	GGCCTCCCTG	CCAACCTCCT	GGCCCTGCGG	GCCTTTGTGG	GGCGGATCCG	CCAGCCCCAG	120
	CCTGCACCTG	TGCACATCCT	CCTGCTGAGC	CTGACGCTGG	CCGACCTCCT	CCTGCTGCTG	180
	CTGCTGCCCT	TCAAGATCAT	CGAGGCTGCG	TCGAACTTCC	GCTGGTACCT	GCCCAAGGTC	240
	GTCTGCGCCC	TCACGAGTTT	TGGCTTCTAC	AGCAGCATCT	ACTGCAGCAC	GTGGCTCCTG	300
	GCGGGCATCA	GCATCGAGCG	CTACCTGGGA	GTGGCTTTCC	CCGTGCAGTA	CAAGCTCTCC	360
30	CGCCGGCCTC	TGTATGGAGT	GATTGCAGCT	CTGGTGGCCT	GGGTTATGTC	CTTTGGTCAC	420
	TGCACCATCG	TGATCATCGT	TCAATACTTG	AACACGACTG	AGCAGGTCAG	AAGTGGCAAT	480
	GAAATTACCT	GCTACGAGAA	CTTCACCGAT	AACCAATTGG	ACGTGGTGCT	GCCCGTGCGG	540
	CTGGAGCTGT	GCCTGGTGCT	CTTCTTCATC	CCCATGGCAG	TCACCATCTT	CTGCTACTGG	600

229

CGTTTTGTGT GGATCATGCT CTCCCAGCCC CTTGTGGGGG CCCAGAGGCG GCGCCGAGCC 660
 AAGGGGCTGG CTGTGGTGAC GCTGCTCAAT TTCCTGGTGT GCTTCGGACC TTACAACGTG 720
 TCCCACCTGG TGGGGTATCA CCAGAGAAAA AGCCCCTGGT GGCGGTCAAT AGCCGTGGTG 780
 TTCAGTTCAC TCAACGCCAG TCTGGACCCC CTGCTCTTCT ATTTCTCTTC TTCAGTGGTG 840
 5 CGCAGGGCAT TTGGGAGAGG GCTGCAGGTG CTGCGGAATC AGGGCTCCTC CCTGTTGGGA 900
 CGCAGAGGCA AAGACACAGC AGAGGGGACA AATGAGGACA GGGGTGTGGG TCAAGGAGAA 960
 GGGATGCCAA GTTCGGAATT CACTACAGAG TAG 993

(277) INFORMATION FOR SEQ ID NO:276:

- 10 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 330 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant
 (ii) MOLECULE TYPE: protein

- 15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:276:

Met Leu Pro Asp Trp Lys Ser Ser Leu Ile Leu Met Ala Tyr Ile Ile
 1 5 10 15
 Ile Phe Leu Thr Gly Leu Pro Ala Asn Leu Leu Ala Leu Arg Ala Phe
 20 25 30
 20 Val Gly Arg Ile Arg Gln Pro Gln Pro Ala Pro Val His Ile Leu Leu
 35 40 45
 Leu Ser Leu Thr Leu Ala Asp Leu Leu Leu Leu Leu Leu Leu Pro Phe
 50 55 60
 25 Lys Ile Ile Glu Ala Ala Ser Asn Phe Arg Trp Tyr Leu Pro Lys Val
 65 70 75 80
 Val Cys Ala Leu Thr Ser Phe Gly Phe Tyr Ser Ser Ile Tyr Cys Ser
 85 90 95
 Thr Trp Leu Leu Ala Gly Ile Ser Ile Glu Arg Tyr Leu Gly Val Ala
 100 105 110
 30 Phe Pro Val Gln Tyr Lys Leu Ser Arg Arg Pro Leu Tyr Gly Val Ile
 115 120 125
 Ala Ala Leu Val Ala Trp Val Met Ser Phe Gly His Cys Thr Ile Val
 130 135 140
 35 Ile Ile Val Gln Tyr Leu Asn Thr Thr Glu Gln Val Arg Ser Gly Asn
 145 150 155 160

Glu Ile Thr Cys Tyr Glu Asn Phe Thr Asp Asn Gln Leu Asp Val Val
 165 170 175
 Leu Pro Val Arg Leu Glu Leu Cys Leu Val Leu Phe Phe Ile Pro Met
 180 185 190
 5 Ala Val Thr Ile Phe Cys Tyr Trp Arg Phe Val Trp Ile Met Leu Ser
 195 200 205
 Gln Pro Leu Val Gly Ala Gln Arg Arg Arg Arg Ala Lys Gly Leu Ala
 210 215 220
 10 Val Val Thr Leu Leu Asn Phe Leu Val Cys Phe Gly Pro Tyr Asn Val
 225 230 235 240
 Ser His Leu Val Gly Tyr His Gln Arg Lys Ser Pro Trp Trp Arg Ser
 245 250 255
 Ile Ala Val Val Phe Ser Ser Leu Asn Ala Ser Leu Asp Pro Leu Leu
 260 265 270
 15 Phe Tyr Phe Ser Ser Ser Val Val Arg Arg Ala Phe Gly Arg Gly Leu
 275 280 285
 Gln Val Leu Arg Asn Gln Gly Ser Ser Leu Leu Gly Arg Arg Gly Lys
 290 295 300
 20 Asp Thr Ala Glu Gly Thr Asn Glu Asp Arg Gly Val Gly Gln Gly Glu
 305 310 315 320
 Gly Met Pro Ser Ser Asp Phe Thr Thr Glu
 325 330

(278) INFORMATION FOR SEQ ID NO:277:

25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2724 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:277:

ATGGACACCT CCCGGCTCGG TGTGCTCCTG TCCTTGCCCTG TGCTGCTGCA GCTGGCGACC 60
 GGGGGCAGCT CTCCCAGGTC TGGTGTGTTG CTGAGGGGCT GCCCCACACA CTGTCATTGC 120
 GAGCCCGACG GCAGGATGTT GCTCAGGGTG GACTGCTCCG ACCTGGGGCT CTCGGAGCTG 180
 CCTTCCAACC TCAGCGTCTT CACCTCCTAC CTAGACCTCA GTATGAACAA CATCAGTCAG 240
 35 CTGCTCCCGA ATCCCCTGCC CAGTCTCCGC TTCCTGGAGG AGTTACGTCT TGCGGGAAAC 300

	GCTCTGACAT	ACATTCCCAA	GGGAGCATTC	ACTGGCCTTT	ACAGTCTTAA	AGTTCTTATG	360
	CTGCAGAATA	ATCAGCTAAG	ACACGTACCC	ACAGAAGCTC	TGCAGAATTT	GCGAAGCCTT	420
	CAATCCCTGC	GTCTGGATGC	TAACCACATC	AGCTATGTGC	CCCCAAGCTG	TTTCAGTGGC	480
	CTGCATTCCC	TGAGGCACCT	GTGGCTGGAT	GACAATGCGT	TAACAGAAAT	CCCCGTCCAG	540
5	GCTTTTAGAA	GTTTATCGGC	ATTGCAAGCC	ATGACCTTGG	CCCTGAACAA	AATACACCAC	600
	ATACCAGACT	ATGCCTTTGG	AAACCTCTCC	AGCTTGGTAG	TTCTACATCT	CCATAACAAT	660
	AGAATCCACT	CCCTGGGAAA	GAAATGCTTT	GATGGGCTCC	ACAGCCTAGA	GACTTTAGAT	720
	TTAAATTACA	ATAACCTTGA	TGAATTCCCC	ACTGCAATTA	GGACACTCTC	CAACCTTAAA	780
	GAACTAGGAT	TTCATAGCAA	CAATATCAGG	TCGATACCTG	AGAAAGCATT	TGTAGGCAAC	840
10	CCTTCTCTTA	TTACAATACA	TTTCTATGAC	AATCCCATCC	AATTTGTTGG	GAGATCTGCT	900
	TTTCAACATT	TACCTGAACT	AAGAACACTG	ACTCTGAATG	GTGCCTCACA	AATAACTGAA	960
	TTTCCTGATT	TAAC TGGAAC	TGCAAACCTG	GAGAGTCTGA	CTTTAACTGG	AGCACAGATC	1020
	TCATCTCTTC	CTCAAACCGT	CTGCAATCAG	TTACCTAATC	TCCAAGTGCT	AGATCTGTCT	1080
	TACAACCTAT	TAGAAGATTT	ACCCAGTTTT	TCAGTCTGCC	AAAAGCTTCA	GAAAATTGAC	1140
15	CTAAGACATA	ATGAAATCTA	CGAAATTAAA	GTTGACACTT	TCCAGCAGTT	GCTTAGCCTC	1200
	CGATCGCTGA	ATTTGGCTTG	GAACAAAATT	GCTATTATTC	ACCCCAATGC	ATTTTCCACT	1260
	TTGCCATCCC	TAATAAAGCT	GGACCTATCG	TCCAACCTCC	TGTCGTCTTT	TCCTATAACT	1320
	GGGTTACATG	GTTTAACTCA	CTTAAAATTA	ACAGGAAATC	ATGCCTTACA	GAGCTTGATA	1380
	TCATCTGAAA	ACTTTCCAGA	ACTCAAGGTT	ATAGAAATGC	CTTATGCTTA	CCAGTGCTGT	1440
20	GCATTTGGAG	TGTGTGAGAA	TGCCTATAAG	ATTTCTAATC	AATGGAATAA	AGGTGACAAC	1500
	AGCAGTATGG	ACGACCTTCA	TAAGAAAGAT	GCTGGAATGT	TTCAGGCTCA	AGATGAACGT	1560
	GACCTTGAAG	ATTTCTTGCT	TGACTTTGAG	GAAGACCTGA	AAGCCCTTCA	TTCAGTGCAG	1620
	TGTTACCTT	CCCCAGGCCC	CTTCAAACCC	TGTGAACACC	TGCTTGATGG	CTGGCTGATC	1680
	AGAATTGGAG	TGTGGACCAT	AGCAGTTCTG	GCACTTACTT	GTAATGCTTT	GGTGACTTCA	1740
25	ACAGTTTTCA	GATCCCCTCT	GTACATTTCC	CCCATTAAAC	TGTTAATTGG	GGTCATCGCA	1800
	GCAGTGAACA	TGCTCACGGG	AGTCTCCAGT	GCCGTGCTGG	CTGGTGTGGA	TGCGTTCACT	1860
	TTTGGCAGCT	TTGCACGACA	TGGTGCCTGG	TGGGAGAATG	GGGTTGGTTG	CCATGTCATT	1920
	GGTTTTTTGT	CCATTTTTTG	TTCAGAATCA	TCTGTTTTCC	TGCTTACTCT	GGCAGCCCTG	1980

GAGCGTGGGT TCTCTGTGAA ATATTCTGCA AAATTGAAA CGAAAGCTCC ATTTTCTAGC 2040
 CTGAAAGTAA TCATTTTGCT CTGTGCCCTG CTGGCCTTGA CCATGGCCGC AGTTCCCCTG 2100
 CTGGGTGGCA GCAAGTATGG CGCCTCCCCT CTCTGCCTGC CTTTGCCTTT TGGGGAGCCC 2160
 AGCACCATGG GCTACATGGT CGCTCTCATC TTGCTCAATT CCCTTTGCTT CCTCATGATG 2220
 5 ACCATTGCCT ACACCAAGCT CTACTGCAAT TTGGACAAGG GAGACCTGGA GAATATTTGG 2280
 GACTGCTCTA TGAAAAACA CATTGCCCTG TTGCTCTTCA CCAACTGCAT CCTAAACTGC 2340
 CCTGTGGCTT TCTTGTCTT CTCCTCTTTA ATAAACCTTA CATTATCAG TCCTGAAGTA 2400
 ATTAAGTTTA TCCTTCTGGT GGTAGTCCCA CTTCTGCAT GTCTCAATCC CCTTCTCTAC 2460
 ATCTTGTTCA ATCCTCACTT TAAGGAGGAT CTGGTGAGCC TGAGAAAGCA AACCTACGTC 2520
 10 TGGACAAGAT CAAAACACCC AAGCTTGATG TCAATTAAC CTGATGATGT CGAAAAACAG 2580
 TCCTGTGACT CAACTCAAGC CTTGGTAACC TTTACCAGCT CCAGCATCAC TTATGACCTG 2640
 CCTCCCAGTT CCGTGCCATC ACCAGCTTAT CCACTGACTG AGAGCTGCCA TCTTCTCTCT 2700
 GTGGCATTG TCCCATGTCT CTAA 2724

(279) INFORMATION FOR SEQ ID NO:278:

- 15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 907 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

- 20 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:278:

Met Asp Thr Ser Arg Leu Gly Val Leu Leu Ser Leu Pro Val Leu Leu
 1 5 10 15
 Gln Leu Ala Thr Gly Gly Ser Ser Pro Arg Ser Gly Val Leu Leu Arg
 25 20 25 30
 Gly Cys Pro Thr His Cys His Cys Glu Pro Asp Gly Arg Met Leu Leu
 35 40 45
 Arg Val Asp Cys Ser Asp Leu Gly Leu Ser Glu Leu Pro Ser Asn Leu
 50 55 60
 30 Ser Val Phe Thr Ser Tyr Leu Asp Leu Ser Met Asn Asn Ile Ser Gln
 65 70 75 80
 Leu Leu Pro Asn Pro Leu Pro Ser Leu Arg Phe Leu Glu Glu Leu Arg
 85 90 95

233

	Leu	Ala	Gly	Asn	Ala	Leu	Thr	Tyr	Ile	Pro	Lys	Gly	Ala	Phe	Thr	Gly	
				100					105					110			
	Leu	Tyr	Ser	Leu	Lys	Val	Leu	Met	Leu	Gln	Asn	Asn	Gln	Leu	Arg	His	
			115					120					125				
5	Val	Pro	Thr	Glu	Ala	Leu	Gln	Asn	Leu	Arg	Ser	Leu	Gln	Ser	Leu	Arg	
		130					135					140					
	Leu	Asp	Ala	Asn	His	Ile	Ser	Tyr	Val	Pro	Pro	Ser	Cys	Phe	Ser	Gly	
	145				150					155						160	
10	Leu	His	Ser	Leu	Arg	His	Leu	Trp	Leu	Asp	Asp	Asn	Ala	Leu	Thr	Glu	
				165					170						175		
	Ile	Pro	Val	Gln	Ala	Phe	Arg	Ser	Leu	Ser	Ala	Leu	Gln	Ala	Met	Thr	
			180						185					190			
	Leu	Ala	Leu	Asn	Lys	Ile	His	His	Ile	Pro	Asp	Tyr	Ala	Phe	Gly	Asn	
			195					200					205				
15	Leu	Ser	Ser	Leu	Val	Val	Leu	His	Leu	His	Asn	Asn	Arg	Ile	His	Ser	
		210					215					220					
	Leu	Gly	Lys	Lys	Cys	Phe	Asp	Gly	Leu	His	Ser	Leu	Glu	Thr	Leu	Asp	
	225				230						235					240	
20	Leu	Asn	Tyr	Asn	Asn	Leu	Asp	Glu	Phe	Pro	Thr	Ala	Ile	Arg	Thr	Leu	
				245						250					255		
	Ser	Asn	Leu	Lys	Glu	Leu	Gly	Phe	His	Ser	Asn	Asn	Ile	Arg	Ser	Ile	
			260						265					270			
	Pro	Glu	Lys	Ala	Phe	Val	Gly	Asn	Pro	Ser	Leu	Ile	Thr	Ile	His	Phe	
			275					280					285				
25	Tyr	Asp	Asn	Pro	Ile	Gln	Phe	Val	Gly	Arg	Ser	Ala	Phe	Gln	His	Leu	
		290				295						300					
	Pro	Glu	Leu	Arg	Thr	Leu	Thr	Leu	Asn	Gly	Ala	Ser	Gln	Ile	Thr	Glu	
	305				310						315					320	
30	Phe	Pro	Asp	Leu	Thr	Gly	Thr	Ala	Asn	Leu	Glu	Ser	Leu	Thr	Leu	Thr	
				325						330					335		
	Gly	Ala	Gln	Ile	Ser	Ser	Leu	Pro	Gln	Thr	Val	Cys	Asn	Gln	Leu	Pro	
				340					345					350			
	Asn	Leu	Gln	Val	Leu	Asp	Leu	Ser	Tyr	Asn	Leu	Leu	Glu	Asp	Leu	Pro	
			355				360						365				
35	Ser	Phe	Ser	Val	Cys	Gln	Lys	Leu	Gln	Lys	Ile	Asp	Leu	Arg	His	Asn	
		370				375						380					
	Glu	Ile	Tyr	Glu	Ile	Lys	Val	Asp	Thr	Phe	Gln	Gln	Leu	Leu	Ser	Leu	

	385		390		395		400									
	Arg	Ser	Leu	Asn	Leu	Ala	Trp	Asn	Lys	Ile	Ala	Ile	Ile	His	Pro	Asn
				405					410							415
5	Ala	Phe	Ser	Thr	Leu	Pro	Ser	Leu	Ile	Lys	Leu	Asp	Leu	Ser	Ser	Asn
				420					425							430
	Leu	Leu	Ser	Ser	Phe	Pro	Ile	Thr	Gly	Leu	His	Gly	Leu	Thr	His	Leu
				435				440					445			
	Lys	Leu	Thr	Gly	Asn	His	Ala	Leu	Gln	Ser	Leu	Ile	Ser	Ser	Glu	Asn
		450				455						460				
10	Phe	Pro	Glu	Leu	Lys	Val	Ile	Glu	Met	Pro	Tyr	Ala	Tyr	Gln	Cys	Cys
	465					470					475					480
	Ala	Phe	Gly	Val	Cys	Glu	Asn	Ala	Tyr	Lys	Ile	Ser	Asn	Gln	Trp	Asn
				485						490						495
15	Lys	Gly	Asp	Asn	Ser	Ser	Met	Asp	Asp	Leu	His	Lys	Lys	Asp	Ala	Gly
				500					505						510	
	Met	Phe	Gln	Ala	Gln	Asp	Glu	Arg	Asp	Leu	Glu	Asp	Phe	Leu	Leu	Asp
			515					520					525			
	Phe	Glu	Glu	Asp	Leu	Lys	Ala	Leu	His	Ser	Val	Gln	Cys	Ser	Pro	Ser
		530				535					540					
20	Pro	Gly	Pro	Phe	Lys	Pro	Cys	Glu	His	Leu	Leu	Asp	Gly	Trp	Leu	Ile
	545					550					555					560
	Arg	Ile	Gly	Val	Trp	Thr	Ile	Ala	Val	Leu	Ala	Leu	Thr	Cys	Asn	Ala
				565						570						575
25	Leu	Val	Thr	Ser	Thr	Val	Phe	Arg	Ser	Pro	Leu	Tyr	Ile	Ser	Pro	Ile
			580					585						590		
	Lys	Leu	Leu	Ile	Gly	Val	Ile	Ala	Ala	Val	Asn	Met	Leu	Thr	Gly	Val
		595					600						605			
	Ser	Ser	Ala	Val	Leu	Ala	Gly	Val	Asp	Ala	Phe	Thr	Phe	Gly	Ser	Phe
		610					615					620				
30	Ala	Arg	His	Gly	Ala	Trp	Trp	Glu	Asn	Gly	Val	Gly	Cys	His	Val	Ile
	625					630					635					640
	Gly	Phe	Leu	Ser	Ile	Phe	Ala	Ser	Glu	Ser	Ser	Val	Phe	Leu	Leu	Thr
				645					650						655	
35	Leu	Ala	Ala	Leu	Glu	Arg	Gly	Phe	Ser	Val	Lys	Tyr	Ser	Ala	Lys	Phe
			660					665						670		
	Glu	Thr	Lys	Ala	Pro	Phe	Ser	Ser	Leu	Lys	Val	Ile	Ile	Leu	Leu	Cys
		675					680						685			

235

Ala Leu Leu Ala Leu Thr Met Ala Ala Val Pro Leu Leu Gly Gly Ser
690 695 700

Lys Tyr Gly Ala Ser Pro Leu Cys Leu Pro Leu Pro Phe Gly Glu Pro
705 710 715 720

5 Ser Thr Met Gly Tyr Met Val Ala Leu Ile Leu Leu Asn Ser Leu Cys
725 730 735

Phe Leu Met Met Thr Ile Ala Tyr Thr Lys Leu Tyr Cys Asn Leu Asp
740 745 750

10 Lys Gly Asp Leu Glu Asn Ile Trp Asp Cys Ser Met Lys Lys His Ile
755 760 765

Ala Leu Leu Leu Phe Thr Asn Cys Ile Leu Asn Cys Pro Val Ala Phe
770 775 780

Leu Ser Phe Ser Ser Leu Ile Asn Leu Thr Phe Ile Ser Pro Glu Val
785 790 795 800

15 Ile Lys Phe Ile Leu Leu Val Val Val Pro Leu Pro Ala Cys Leu Asn
805 810 815

Pro Leu Leu Tyr Ile Leu Phe Asn Pro His Phe Lys Glu Asp Leu Val
820 825 830

20 Ser Leu Arg Lys Gln Thr Tyr Val Trp Thr Arg Ser Lys His Pro Ser
835 840 845

Leu Met Ser Ile Asn Ser Asp Asp Val Glu Lys Gln Ser Cys Asp Ser
850 855 860

Thr Gln Ala Leu Val Thr Phe Thr Ser Ser Ser Ile Thr Tyr Asp Leu
865 870 875 880

25 Pro Pro Ser Ser Val Pro Ser Pro Ala Tyr Pro Val Thr Glu Ser Cys
885 890 895

His Leu Ser Ser Val Ala Phe Val Pro Cys Leu
900 905

(280) INFORMATION FOR SEQ ID NO:279:

30

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 32 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

35

- (ii) MOLECULE TYPE: DNA (genomic)

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:279:

CATGCCAACC GGCCGCGAG GCTGCTGCTG GT

(281) INFORMATION FOR SEQ ID NO:280:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

5

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:280:

ACCAGCAGCA GCCTCGCGGG CCGGTTGGCA TG

32

INTERNATIONAL SEARCH REPORT

Int. Patent Application No.

PCT/US 99/23938

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/12 C07K14/72 G01N33/50 G01N33/566

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C07K G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	KJELSBERG M. A. ET AL.: "CONSTITUTIVE ACTIVATION OF THE ALPHA1B-ADRENERGIC RECEPTOR BY ALL AMINO ACID SUBSTITUTIONS AT A SINGLE SITE" JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 267, no. 3, 25 January 1992 (1992-01-25), pages 1430-1433, XP002911764 ISSN: 0021-9258 the whole document ---	1,2, 4-13, 15-33, 35-37,41

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

2 March 2000

Date of mailing of the international search report

09/03/2000

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Mandl, B

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SCHEER A. ET AL.: "CONSTITUTIVELY ACTIVE G PROTEIN-COUPLED RECEPTORS: POTENTIAL MECHANISMS OF RECEPTOR ACTIVATION" JOURNAL OF RECEPTOR AND SIGNAL TRANSDUCTION RESEARCH, vol. 17, no. 1/03, 1997, pages 57-73, XP000867531 ISSN: 1079-9893 the whole document	1,2, 4-13, 15-33, 35-37,41
X	WO 97 21731 A (NEW ENGLAND MEDICAL CENTER INC) 19 June 1997 (1997-06-19) the whole document, especially Fig. 2-3	1,2,4, 9-13, 20-32, 35-37,41
X	WO 98 38217 A (HERRICK DAVIS KATHARINE ;TEITLER MILT (US); EGAN CHRISTINA C (US)) 3 September 1998 (1998-09-03) the whole document, especially page 7, lines 24-27, and figure 4	1,2, 4-13, 15-33, 35-37,41
P,X	PAUWELS P. J. ET AL.: "REVIEW: AMINO ACID DOMAINS INVOLVED IN CONSTITUTIVE ACTIVATION OF G-PROTEIN-COUPLED RECEPTORS" MOLECULAR NEUROBIOLOGY, vol. 17, no. 1/03, 1998, pages 109-135, XP000866477 ISSN: 0893-7648 the whole document	1,2, 4-13, 15-33, 35-37,41

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 99/ 23938

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☒ Claims Nos.: 34, 38-40, 42, 43
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
See FURTHER INFORMATION sheet PCT/ISA/210
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US 99 23938

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 34,38-40,42,43

Claims 34, 38-40, 42 and 43 refer to compounds with an agonistic effect on a GPCR without giving a true technical characterization. Moreover, no such specific compounds are defined in the application. In consequence, the scope of said claims is ambiguous and vague, and their subject-matter is not sufficiently disclosed and supported (Art. 5 and 6 PCT). No search can be carried out for such purely speculative claims whose wording is, in fact, a mere recitation of the results to be achieved.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 99/23938

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9721731 A	19-06-1997	US 5750353 A	12-05-1998
		AU 1334397 A	03-07-1997
		CA 2239293 A	19-06-1997
		EP 0869975 A	14-10-1998
WO 9838217 A	03-09-1998	AU 6343998 A	18-09-1998



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : C12N 15/12, C07K 14/72, G01N 33/50, 33/566	A1	(11) International Publication Number: WO 00/22129 (43) International Publication Date: 20 April 2000 (20.04.00)
(21) International Application Number: PCT/US99/23938 (22) International Filing Date: 12 October 1999 (12.10.99) (30) Priority Data: 09/170,496 13 October 1998 (13.10.98) US (63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application US 09/170,496 (CIP) Filed on 13 October 1998 (13.10.98) (71) Applicant (for all designated States except US): ARENA PHARMACEUTICALS, INC. [US/US]; 6166 Nancy Ridge Drive, San Diego, CA 92121 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): BEHAN, Dominic, P. [GB/US]; 11472 Roxboro Court, San Diego, CA 92131 (US). CHALMERS, Derek, T. [GB/US]; 347 Longden Lane, Solana Beach, CA 92075 (US). IIAW, Chen, W. [US/US]; 7668 Salix Place, San Diego, CA 92129 (US).		(74) Agents: MILLER, Suzanne, E. et al.; Woodcock Washburn Kurtz Mackiewicz & Norris LLP, 46th floor, One Liberty Place, Philadelphia, PA 19103 (US). (81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
(54) Title: NON-ENDOGENOUS, CONSTITUTIVELY ACTIVATED HUMAN G PROTEIN-COUPLED RECEPTORS		
(57) Abstract <p>Disclosed herein are constitutively activated, non-endogenous versions of endogenous human G protein-coupled receptors comprising (a) the following amino acid sequence region (C-terminus to N-terminus orientation) and/or (b) the following nucleic acid sequence region (3' to 5' orientation) transverse the transmembrane-6 (TM6) and intracellular loop-3 (IC3) regions of the GPCR: (a) P¹ AA₁₅ X and/or (b) P^{codon} (AA-codon)₁₅ X_{codon}, respectively. In a most preferred embodiment, P¹ and P^{codon} are endogenous proline and an endogenous nucleic acid encoding region encoding proline, respectively, located within TM6 of the non-endogenous GPCR; AA₁₅ and (AA-codon)₁₅ are 15 endogenous amino acid residues and 15 codons encoding endogenous amino acid residues, respectively; and X and X_{codon} are non-endogenous lysine and a non-endogenous nucleic acid encoding region encoding lysine, respectively, located within IC3 of the non-endogenous GPCR. Because it is most preferred that the non-endogenous human GPCRs which incorporate these mutations are incorporated into mammalian cells and utilized for the screening of the candidate compounds, the non-endogenous human GPCR incorporating the mutation need not be purified and isolated <i>per se</i> (i.e., these are incorporated within the cellular membrane of a mammalian cell), although such purified and isolated non-endogenous human GPCRs are well within the purview of this disclosure.</p>		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

**NON-ENDOGENOUS, CONSTITUTIVELY ACTIVATED
HUMAN G PROTEIN-COUPLED RECEPTORS**

The benefits of commonly owned U.S. Serial Number 09/170,496, filed October 13, 1998, U.S. Serial Number 08/839, 449 filed April 14, 1997 (now abandoned),
5 U.S. Serial Number 09/060,188, filed April 14, 1998; U.S. Provisional Number 60/090,783, filed June 26, 1998; and U.S. Provisional Number 60/095,677, filed on August 7, 1998, are hereby claimed. Each of the foregoing applications are incorporated by reference herein in their entirety.

FIELD OF THE INVENTION

10 The invention disclosed in this patent document relates to transmembrane receptors, and more particularly to human G protein-coupled receptors (GPCRs) which have been altered such that altered GPCRs are constitutively activated. Most preferably, the altered human GPCRs are used for the screening of therapeutic compounds.

BACKGROUND OF THE INVENTION

Although a number of receptor classes exist in humans, by far the most abundant and therapeutically relevant is represented by the G protein-coupled receptor (GPCR or GPCRs) class.

It is estimated that there are some 100,000 genes within the human genome, and of these,

5 approximately 2% or 2,000 genes, are estimated to code for GPCRs. Of these, there are approximately 100 GPCRs for which the endogenous ligand that binds to the GPCR has been identified. Because of the significant time-lag that exists between the discovery of an endogenous GPCR and its endogenous ligand, it can be presumed that the remaining 1,900 GPCRs will be identified and characterized long before the endogenous ligands for these receptors are identified.

10 Indeed, the rapidity by which the Human Genome Project is sequencing the 100,000 human genes indicates that the remaining human GPCRs will be fully sequenced within the next few years. Nevertheless, and despite the efforts to sequence the human genome, it is still very unclear as to how scientists will be able to rapidly, effectively and efficiently exploit this information to improve and enhance the human condition. The present invention is geared towards this
15 important objective.

Receptors, including GPCRs, for which the endogenous ligand has been identified are referred to as "known" receptors, while receptors for which the endogenous ligand has not been identified are referred to as "orphan" receptors. This distinction is not merely semantic, particularly in the case of GPCRs. GPCRs represent an important area for the development of
20 pharmaceutical products: from approximately 20 of the 100 known GPCRs, 60% of all prescription pharmaceuticals have been developed. Thus, the orphan GPCRs are to the pharmaceutical industry what gold was to California in the late 19th century – an opportunity to drive growth, expansion, enhancement and development. A serious drawback exists, however,

- 3 -

with orphan receptors relative to the discovery of novel therapeutics. This is because the traditional approach to the discovery and development of pharmaceuticals has required access to both the receptor *and* its endogenous ligand. Thus, heretofore, orphan GPCRs have presented the art with a tantalizing and undeveloped resource for the discovery of pharmaceuticals.

5 Under the traditional approach to the discovery of potential therapeutics, it is generally the case that the receptor is first identified. Before drug discovery efforts can be initiated, elaborate, time consuming and expensive procedures are typically put into place in order to identify, isolate and generate the receptor's endogenous ligand – this process can require from between 3 and ten years per receptor, at a cost of about \$5million (U.S.) per receptor. These time and financial
10 resources must be expended before the traditional approach to drug discovery can commence. This is because traditional drug discovery techniques rely upon so-called "competitive binding assays" whereby putative therapeutic agents are "screened" against the receptor in an effort to discover compounds that either block the endogenous ligand from binding to the receptor ("antagonists"), or enhance or mimic the effects of the ligand binding to the receptor ("agonists").
15 The overall objective is to identify compounds that prevent cellular activation when the ligand binds to the receptor (the antagonists), or that enhance or increase cellular activity that would otherwise occur if the ligand was properly binding with the receptor (the agonists). Because the endogenous ligands for orphan GPCRs are by definition not identified, the ability to discover novel and unique therapeutics to these receptors using traditional drug discovery techniques is not
20 possible. The present invention, as will be set forth in greater detail below, overcomes these and other severe limitations created by such traditional drug discovery techniques.

GPCRs share a common structural motif. All these receptors have seven sequences of between 22 to 24 hydrophobic amino acids that form seven alpha helices, each of which spans the

membrane (each span is identified by number, *i.e.*, transmembrane-1 (TM-1), transmembrane-2 (TM-2), etc.). The transmembrane helices are joined by strands of amino acids between transmembrane-2 and transmembrane-3, transmembrane-4 and transmembrane-5, and transmembrane-6 and transmembrane-7 on the exterior, or "extracellular" side, of the cell membrane (these are referred to as "extracellular" regions 1, 2 and 3 (EC-1, EC-2 and EC-3), respectively). The transmembrane helices are also joined by strands of amino acids between transmembrane-1 and transmembrane-2, transmembrane-3 and transmembrane-4, and transmembrane-5 and transmembrane-6 on the interior, or "intracellular" side, of the cell membrane (these are referred to as "intracellular" regions 1, 2 and 3 (IC-1, IC-2 and IC-3), respectively). The "carboxy" ("C") terminus of the receptor lies in the intracellular space within the cell, and the "amino" ("N") terminus of the receptor lies in the extracellular space outside of the cell. The general structure of G protein-coupled receptors is depicted in Figure 1.

Generally, when an endogenous ligand binds with the receptor (often referred to as "activation" of the receptor), there is a change in the conformation of the intracellular region that allows for coupling between the intracellular region and an intracellular "G-protein." Although other G proteins exist, currently, Gq, Gs, Gi, and Go are G proteins that have been identified. Endogenous ligand-activated GPCR coupling with the G-protein begins a signaling cascade process (referred to as "signal transduction"). Under normal conditions, signal transduction ultimately results in cellular activation or cellular inhibition. It is thought that the IC-3 loop as well as the carboxy terminus of the receptor interact with the G protein. A principal focus of this invention is directed to the transmembrane-6 (TM6) region and the intracellular-3 (IC3) region of the GPCR.

Under physiological conditions, GPCRs exist in the cell membrane in equilibrium between

- 5 -

two different conformations: an "inactive" state and an "active" state. As shown schematically in Figure 2, a receptor in an inactive state is unable to link to the intracellular signaling transduction pathway to produce a biological response. Changing the receptor conformation to the active state allows linkage to the transduction pathway (via the G-protein) and produces a biological response.

5 A receptor may be stabilized in an active state by an endogenous ligand or a compound such as a drug. Recent discoveries, including but not exclusively limited to modifications to the amino acid sequence of the receptor, provide means other than endogenous ligands or drugs to promote and stabilize the receptor in the active state conformation. These means effectively stabilize the receptor in an active state by simulating the effect of an endogenous ligand binding
10 to the receptor. Stabilization by such ligand-independent means is termed "constitutive receptor activation."

As noted above, the use of an orphan receptor for screening purposes has not been possible. This is because the traditional "dogma" regarding screening of compounds mandates that the ligand for the receptor be known. By definition, then, this approach has no applicability with
15 respect to orphan receptors. Thus, by adhering to this dogmatic approach to the discovery of therapeutics, the art, in essence, has taught and has been taught to forsake the use of orphan receptors unless and until the endogenous ligand for the receptor is discovered. Given that there are an estimated 2,000 G protein coupled receptors, the majority of which are orphan receptors, such dogma castigates a creative, unique and distinct approach to the discovery of therapeutics.

20 Information regarding the nucleic acid and/or amino acid sequences of a variety of GPCRs is summarized below in Table A. Because an important focus of the invention disclosed herein is directed towards orphan GPCRs, many of the below-cited references are related to orphan GPCRs. However, this list is not intended to imply, nor is this list to be construed, legally or

otherwise, that the invention disclosed herein is only applicable to orphan GPCRs or the specific GPCRs listed below. Additionally, certain receptors that have been isolated are not the subject of publications per se; for example, reference is made to a G Protein-Coupled Receptor database on the "world-wide web" (neither the named inventors nor the assignee have any affiliation with this site) that lists GPCRs. Other GPCRs are the subject of patent applications owned by the present assignee and these are not listed below (including GPR3, GPR6 and GPR12; see U.S. Provisional Number 60/094879):

Table A

Receptor Name	Publication Reference
GPR1	23 Genomics 609 (1994)
GPR4	14 DNA and Cell Biology 25 (1995)
GPR5	14 DNA and Cell Biology 25 (1995)
GPR7	28 Genomics 84 (1995)
GPR8	28 Genomics 84 (1995)
GPR9	184 J. Exp. Med. 963 (1996)
GPR10	29 Genomics 335 (1995)
GPR15	32 Genomics 462 (1996)
GPR17	70 J Neurochem. 1357 (1998)
GPR18	42 Genomics 462 (1997)
GPR20	187 Gene 75 (1997)
GPR21	187 Gene 75 (1997)
GPR22	187 Gene 75 (1997)
GPR24	398 FEBS Lett. 253 (1996)
GPR30	45 Genomics 607 (1997)
GPR31	42 Genomics 519 (1997)
GPR32	50 Genomics 281 (1997)
GPR40	239 Biochem. Biophys. Res. Commun. 543 (1997)
GPR41	239 Biochem. Biophys. Res. Commun. 543 (1997)
GPR43	239 Biochem. Biophys. Res. Commun. 543 (1997)
APJ	136 Gene 355 (1993)
BLR1	22 Eur. J. Immunol. 2759 (1992)
CEPR	231 Biochem. Biophys. Res. Commun. 651 (1997)
EBI1	23 Genomics 643 (1994)
EBI2	67 J. Virol. 2209 (1993)
ETBR-LP2	424 FEBS Lett. 193 (1998)
GPCR-CNS	54 Brain Res. Mol. Brain Res. 152 (1998); 45 Genomics 68 (1997)
GPR-NGA	394 FEBS Lett. 325 (1996)
H9	386 FEBS Lett 219 (1996)

- 7 -

HBA954	1261 Biochim. Biophys. Acta 121 (1995)
HG38	247 Biochem. Biophys. Res. Commun. 266 (1998)
HM74	5 Int. Immunol. 1239 (1993)
OGR1	35 Genomics 397 (1996)
V28	163 Gene 295 (1995)

5

As will be set forth and disclosed in greater detail below, utilization of a mutational cassette to modify the endogenous sequence of a human GPCR leads to a constitutively activated version of the human GPCR. These non-endogenous, constitutively activated versions of human GPCRs can be utilized, *inter alia*, for the screening of candidate compounds to directly identify compounds of, *e.g.*, therapeutic relevance.

10

SUMMARY OF THE INVENTION

Disclosed herein is a non-endogenous, human G protein-coupled receptor comprising (a) as a most preferred amino acid sequence region (C-terminus to N-terminus orientation) and/or (b) as a most preferred nucleic acid sequence region (3' to 5' orientation) transversing the transmembrane-6 (TM6) and intracellular loop-3 (IC3) regions of the GPCR:

15

(a) P¹ AA₁₅ X

wherein:

- (1) P¹ is an amino acid residue located within the TM6 region of the GPCR, where P¹ is selected from the group consisting of (i) the endogenous GPCR's proline residue, and (ii) a non-endogenous amino acid residue other than proline;
- (2) AA₁₅ are 15 amino acids selected from the group consisting of

20

5 (a) the endogenous GPCR's amino acids (b) non-endogenous amino acid residues, and (c) a combination of the endogenous GPCR's amino acids and non-endogenous amino acids, excepting that none of the 15 endogenous amino acid residues that are positioned within the TM6 region of the GPCR is proline; and

10 (3) X is a non-endogenous amino acid residue located within the IC3 region of said GPCR, preferably selected from the group consisting of lysine, histidine and arginine, and most preferably lysine, excepting that when the endogenous amino acid at position X is lysine, then X is an amino acid other than lysine, preferably alanine;

and/or

(b) P^{codon} (AA-codon)₁₅ X_{codon}

15 wherein:

- (1) P^{codon} is a nucleic acid sequence within the TM6 region of the GPCR, where P^{codon} encodes an amino acid selected from the group consisting of (i) the endogenous GPCR's proline residue, and (ii) a non-endogenous amino acid residue other than proline;
- 20 (2) (AA-codon)₁₅ are 15 codons encoding 15 amino acids selected from the group consisting of (a) the endogenous GPCR's amino acids (b) non-endogenous amino acid residues and (c) a combination of the endogenous GPCR's amino acids and non-

- 9 -

endogenous amino acids, excepting that none of the 15 endogenous codons within the TM6 region of the GPCR encodes a proline amino acid residue; and

- (3) X_{codon} is a nucleic acid encoding region residue located within the IC3 region of said GPCR, where X_{codon} encodes a non-endogenous amino acid, preferably selected from the group consisting of lysine, histidine and arginine, and most preferably lysine, excepting that when the endogenous encoding region at position X_{codon} encodes the amino acid lysine, then X_{codon} encodes an amino acid other than lysine, preferably alanine.

The terms endogenous and non-endogenous in reference to these sequence cassettes are relative to the endogenous GPCR. For example, once the endogenous proline residue is located within the TM6 region of a particular GPCR, and the 16th amino acid therefrom is identified for mutation to constitutively activate the receptor, it is also possible to mutate the endogenous proline residue (*i.e.*, once the marker is located and the 16th amino acid to be mutated is identified, one may mutate the marker itself), although it is most preferred that the proline residue not be mutated. Similarly, and while it is most preferred that AA₁₅ be maintained in their endogenous forms, these amino acids may also be mutated. The only amino acid that must be mutated in the non-endogenous version of the human GPCR is X *i.e.*, the endogenous amino acid that is 16 residues from P¹ cannot be maintained in its endogenous form and must be mutated, as further disclosed herein. Stated again, while it is preferred that in the non-endogenous version of the human GPCR, P¹ and AA₁₅ remain in their endogenous forms (*i.e.*, identical to their wild-type forms), once X is identified and mutated, any and/or all of P¹ and AA₁₅ can be mutated. This applies to the nucleic

acid sequences as well. In those cases where the endogenous amino acid at position X is lysine, then in the non-endogenous version of such GPCR, X is an amino acid other than lysine, preferably alanine.

Accordingly, and as a hypothetical example, if the endogenous GPCR has the following
5 endogenous amino acid sequence at the above-noted positions:

P-AACCTTGRRRDDDE -Q

then any of the following exemplary and hypothetical cassettes would fall within the scope of the disclosure (non-endogenous amino acids are set forth in bold):

P-AACCTTGRRRDDDE -K

10 P-AACCTTHIGRRRDDDE -K

P-ADEETTGGRRRDDDE -A

P-LLKFMSTWZLVAAPQ -K

A-LLKFMSTWZLVAAPQ -K

It is also possible to add amino acid residues within AA₁₅, but such an approach is not particularly
15 advanced. Indeed, in the most preferred embodiments, the only amino acid that differs in the non-endogenous version of the human GPCR as compared with the endogenous version of that GPCR is the amino acid in position X; mutation of this amino acid itself leads to constitutive activation of the receptor.

Thus, in particularly preferred embodiments, P¹ and P^{codon} are endogenous proline and an
20 endogenous nucleic acid encoding region encoding proline, respectively; and X and X_{codon} are non-endogenous lysine or alanine and a non-endogenous nucleic acid encoding region encoding lysine or alanine, respectively, with lysine being most preferred. Because it is most preferred that the non-endogenous versions of the human GPCRs which incorporate these mutations are

- 11 -

incorporated into mammalian cells and utilized for the screening of candidate compounds, the non-endogenous human GPCR incorporating the mutation need not be purified and isolated *per se* (i.e., these are incorporated within the cellular membrane of a mammalian cell), although such purified and isolated non-endogenous human GPCRs are well within the purview of this disclosure. Gene-
5 targeted and transgenic non-human mammals (preferably rats and mice) incorporating the non-endogenous human GPCRs are also within the purview of this invention; in particular, gene-targeted mammals are most preferred in that these animals will incorporate the non-endogenous versions of the human GPCRs in place of the non-human mammal's endogenous GPCR-encoding region (techniques for generating such non-human mammals to replace the non-human mammal's
10 protein encoding region with a human encoding region are well known; see, for example, U.S. Patent No. 5,777,194.)

It has been discovered that these changes to an endogenous human GPCR render the GPCR constitutively active such that, as will be further disclosed herein, the non-endogenous, constitutively activated version of the human GPCR can be utilized for, *inter alia*, the direct
15 screening of candidate compounds without the need for the endogenous ligand. Thus, methods for using these materials, and products identified by these methods are also within the purview of the following disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a generalized structure of a G protein-coupled receptor with the numbers
20 assigned to the transmembrane helices, the intracellular loops, and the extracellular loops.

Figure 2 schematically shows the two states, active and inactive, for a typical G protein coupled receptor and the linkage of the active state to the second messenger transduction pathway.

Figure 3 is a sequence diagram of the preferred vector pCMV, including restriction enzymes site locations.

Figure 4 is a diagrammatic representation of the signal measured comparing pCMV, non-endogenous, constitutively active GPR30 inhibition of GPR6-mediated activation of CRE-Luc reporter with endogenous GPR30 inhibition of GPR6-mediated activation of CRE-Luc reporter.

Figure 5 is a diagrammatic representation of the signal measured comparing pCMV, non-endogenous, constitutively activated GPR17 inhibition of GPR3-mediated activation of CRE-Luc reporter with endogenous GPR17 inhibition of GPR3-mediated activation of CRE-Luc reporter.

Figure 6 provides diagrammatic results of the signal measured comparing control pCMV, endogenous APJ and non-endogenous APJ.

Figure 7 provides an illustration of IP₃ production from non-endogenous human 5-HT_{2A} receptor as compared to the endogenous version of this receptor.

Figure 8 are dot-blot format results for GPR1 (8A), GPR30 (8B) and APJ (8C).

DETAILED DESCRIPTION

The scientific literature that has evolved around receptors has adopted a number of terms to refer to ligands having various effects on receptors. For clarity and consistency, the following definitions will be used throughout this patent document. To the extent that these definitions conflict with other definitions for these terms, the following definitions shall control:

AGONISTS shall mean compounds that activate the intracellular response when they bind to the receptor, or enhance GTP binding to membranes.

- 13 -

AMINO ACID ABBREVIATIONS used herein are set below:

	ALANINE	ALA	A
	ARGININE	ARG	R
	ASPARAGINE	ASN	N
5	ASPARTIC ACID	ASP	D
	CYSTEINE	CYS	C
	GLUTAMIC ACID	GLU	E
	GLUTAMINE	GLN	Q
	GLYCINE	GLY	G
10	HISTIDINE	HIS	H
	ISOLEUCINE	ILE	I
	LEUCINE	LEU	L
	LYSINE	LYS	K
	METHIONINE	MET	M
15	PHENYLALANINE	PHE	F
	PROLINE	PRO	P
	SERINE	SER	S
	THREONINE	THR	T
	TRYPTOPHAN	TRP	W
20	TYROSINE	TYR	Y
	VALINE	VAL	V

PARTIAL AGONISTS shall mean compounds which activate the intracellular response when they bind to the receptor to a lesser degree/extent than do agonists, or enhance GTP binding to membranes to a lesser degree/extent than do agonists

25 ANTAGONIST shall mean compounds that competitively bind to the receptor at the same site as the agonists but which do not activate the intracellular response initiated by the active form of the receptor, and can thereby inhibit the intracellular responses by agonists or partial agonists. ANTAGONISTS do not diminish the baseline intracellular response in the absence of an agonist or partial agonist.

30 CANDIDATE COMPOUND shall mean a molecule (for example, and not limitation, a chemical compound) which is amenable to a screening technique. Preferably, the phrase

"candidate compound" does not include compounds which were publicly known to be compounds selected from the group consisting of inverse agonist, agonist or antagonist to a receptor, as previously determined by an indirect identification process ("indirectly identified compound"); more preferably, not including an indirectly identified compound which has previously been
5 determined to have therapeutic efficacy in at least one mammal; and, most preferably, not including an indirectly identified compound which has previously been determined to have therapeutic utility in humans.

CODON shall mean a grouping of three nucleotides (or equivalents to nucleotides) which generally comprise a nucleoside (adenosine (A), guanosine (G), cytidine (C), uridine (U) and
10 thymidine (T)) coupled to a phosphate group and which, when translated, encodes an amino acid.

COMPOUND EFFICACY shall mean a measurement of the ability of a compound to inhibit or stimulate receptor functionality, as opposed to receptor binding affinity. A preferred means of detecting compound efficacy is via measurement of, *e.g.*, [³⁵S]GTPγS binding, as further disclosed in the Example section of this patent document.

15 CONSTITUTIVELY ACTIVATED RECEPTOR shall mean a receptor subject to constitutive receptor activation. In accordance with the invention disclosed herein, a non-endogenous, human constitutively activated G protein-coupled receptor is one that has been mutated to include the amino acid cassette P¹AA₁₅X, as set forth in greater detail below.

CONSTITUTIVE RECEPTOR ACTIVATION shall mean stabilization of a receptor
20 in the active state by means other than binding of the receptor with its endogenous ligand or a chemical equivalent thereof. Preferably, a G protein-coupled receptor subjected to constitutive receptor activation in accordance with the invention disclosed herein evidences at least a 10% difference in response (increase or decrease, as the case may be) to the signal measured for

- 15 -

constitutive activation as compared with the endogenous form of that GPCR, more preferably, about a 25% difference in such comparative response, and most preferably about a 50% difference in such comparative response. When used for the purposes of directly identifying candidate compounds, it is most preferred that the signal difference be at least about 50% such that there is
5 a sufficient difference between the endogenous signal and the non-endogenous signal to differentiate between selected candidate compounds. In most instances, the "difference" will be an increase in signal; however, with respect to Gs-coupled GPCRS, the "difference" measured is preferably a decrease, as will be set forth in greater detail below.

CONTACT or **CONTACTING** shall mean bringing at least two moieties together,
10 whether in an in vitro system or an in vivo system.

DIRECTLY IDENTIFYING or **DIRECTLY IDENTIFIED**, in relationship to the phrase "candidate compound", shall mean the screening of a candidate compound against a constitutively activated G protein-coupled receptor, and assessing the compound efficacy of such compound. This phrase is, under no circumstances, to be interpreted or understood to be
15 encompassed by or to encompass the phrase "indirectly identifying" or "indirectly identified."

ENDOGENOUS shall mean a material that is naturally produced by the genome of the species. **ENDOGENOUS** in reference to, for example and not limitation, GPCR, shall mean that which is naturally produced by a human, an insect, a plant, a bacterium, or a virus. By contrast, the term **NON-ENDOGENOUS** in this context shall mean that which is not naturally produced
20 by the genome of a species. For example, and not limitation, a receptor which is not constitutively active in its endogenous form, but when mutated by using the cassettes disclosed herein and thereafter becomes constitutively active, is most preferably referred to herein as a "non-endogenous, constitutively activated receptor." Both terms can be utilized to describe both "in

vivo" and "in vitro" systems. For example, and not limitation, in a screening approach, the endogenous or non-endogenous receptor may be in reference to an in vitro screening system whereby the receptor is expressed on the cell-surface of a mammalian cell. As a further example and not limitation, where the genome of a mammal has been manipulated to include a non-
5 endogenous constitutively activated receptor, screening of a candidate compound by means of an in vivo system is viable.

HOST CELL shall mean a cell capable of having a Plasmid and/or Vector incorporated therein. In the case of a prokaryotic Host Cell, a Plasmid is typically replicated as an autonomous molecule as the Host Cell replicates (generally, the Plasmid is thereafter isolated for introduction
10 into a eukaryotic Host Cell); in the case of a eukaryotic Host Cell, a Plasmid is integrated into the cellular DNA of the Host Cell such that when the eukaryotic Host Cell replicates, the Plasmid replicates. Preferably, for the purposes of the invention disclosed herein, the Host Cell is eukaryotic, more preferably, mammalian, and most preferably selected from the group consisting of 293, 293T and COS-7 cells.

15 INDIRECTLY IDENTIFYING or INDIRECTLY IDENTIFIED means the traditional approach to the drug discovery process involving identification of an endogenous ligand specific for an endogenous receptor, screening of candidate compounds against the receptor for determination of those which interfere and/or compete with the ligand-receptor interaction, and assessing the efficacy of the compound for affecting at least one second messenger pathway
20 associated with the activated receptor.

INHIBIT or INHIBITING, in relationship to the term "response" shall mean that a response is decreased or prevented in the presence of a compound as opposed to in the absence of
und.

- 17 -

INVERSE AGONISTS shall mean compounds which bind to either the endogenous form of the receptor or to the constitutively activated form of the receptor, and which inhibit the baseline intracellular response initiated by the active form of the receptor below the normal base level of activity which is observed in the absence of agonists or partial agonists, or decrease GTP binding to membranes. Preferably, the baseline intracellular response is inhibited in the presence of the inverse agonist by at least 30%, more preferably by at least 50%, and most preferably by at least 75%, as compared with the baseline response in the absence of the inverse agonist.

KNOWN RECEPTOR shall mean an endogenous receptor for which the endogenous ligand specific for that receptor has been identified.

10 **LIGAND** shall mean an endogenous, naturally occurring molecule specific for an endogenous, naturally occurring receptor.

MUTANT or **MUTATION** in reference to an endogenous receptor's nucleic acid and/or amino acid sequence shall mean a specified change or changes to such endogenous sequences such that a mutated form of an endogenous, non-constitutively activated receptor evidences constitutive activation of the receptor. In terms of equivalents to specific sequences, a subsequent mutated form of a human receptor is considered to be equivalent to a first mutation of the human receptor if (a) the level of constitutive activation of the subsequent mutated form of the receptor is substantially the same as that evidenced by the first mutation of the receptor; and (b) the percent sequence (amino acid and/or nucleic acid) homology between the subsequent mutated form of the receptor and the first mutation of the receptor is at least about 80%, more preferably at least about 90% and most preferably at least 95%. Ideally, and owing to the fact that the most preferred cassettes disclosed herein for achieving constitutive activation includes a single amino acid and/or codon change between the endogenous and the non-endogenous forms of the GPCR (i.e. X or

X_{codon}), the percent sequence homology should be at least 98%.

ORPHAN RECEPTOR shall mean an endogenous receptor for which the endogenous ligand specific for that receptor has not been identified or is not known.

PHARMACEUTICAL COMPOSITION shall mean a composition comprising at least
5 one active ingredient, whereby the composition is amenable to investigation for a specified, efficacious outcome in a mammal (for example, and not limitation, a human). Those of ordinary skill in the art will understand and appreciate the techniques appropriate for determining whether an active ingredient has a desired efficacious outcome based upon the needs of the artisan.

PLASMID shall mean the combination of a Vector and cDNA. Generally, a Plasmid is
10 introduced into a Host Cell for the purpose of replication and/or expression of the cDNA as a protein.

STIMULATE or **STIMULATING**, in relationship to the term "response" shall mean that a response is increased in the presence of a compound as opposed to in the absence of the compound.

15 **TRANSVERSE** or **TRANSVERSING**, in reference to either a defined nucleic acid sequence or a defined amino acid sequence, shall mean that the sequence is located within at least two different and defined regions. For example, in an amino acid sequence that is 10 amino acid moieties in length, where 3 of the 10 moieties are in the TM6 region of a GPCR and the remaining 7 moieties are in the IC3 region of the GPCR, the 10 amino acid moiety can be described as
20 transversing the TM6 and IC3 regions of the GPCR.

VECTOR in reference to cDNA shall mean a circular DNA capable of incorporating at least one cDNA and capable of incorporation into a Host Cell.

The order of the following sections is set forth for presentational efficiency and is not

- 19 -

intended, nor should be construed, as a limitation on the disclosure or the claims to follow:

A. Introduction

The traditional study of receptors has always proceeded from the a priori assumption (historically based) that the endogenous ligand must first be identified before discovery could proceed to find antagonists and other molecules that could affect the receptor. Even in cases where an antagonist might have been known first, the search immediately extended to looking for the endogenous ligand. This mode of thinking has persisted in receptor research even after the discovery of constitutively activated receptors. What has not been heretofore recognized is that it is the active state of the receptor that is most useful for discovering agonists, partial agonists, and inverse agonists of the receptor. For those diseases which result from an overly active receptor or an under-active receptor, what is desired in a therapeutic drug is a compound which acts to diminish the active state of a receptor or enhance the activity of the receptor, respectively, not necessarily a drug which is an antagonist to the endogenous ligand. This is because a compound that reduces or enhances the activity of the active receptor state need not bind at the same site as the endogenous ligand. Thus, as taught by a method of this invention, any search for therapeutic compounds should start by screening compounds against the ligand-independent active state.

Screening candidate compounds against non-endogenous, constitutively activated GPCRs allows for the direct identification of candidate compounds which act at these cell surface receptors, without requiring any prior knowledge or use of the receptor's endogenous ligand. By determining areas within the body where the endogenous version of such GPCRs are expressed and/or over-expressed, it is possible to determine related disease/disorder states which are associated with the expression and/or over-expression of these receptors; such an approach is disclosed in this patent document.

B. Disease/Disorder Identification and/ r Selection

Most preferably, inverse agonists to the non-endogenous, constitutively activated GPCRs can be identified using the materials of this invention. Such inverse agonists are ideal candidates as lead compounds in drug discovery programs for treating diseases related to these receptors.

- 5 Because of the ability to directly identify inverse agonists, partial agonists or agonists to these receptors, thereby allowing for the development of pharmaceutical compositions, a search, for diseases and disorders associated with these receptors is possible. For example, scanning both diseased and normal tissue samples for the presence of these receptor now becomes more than an academic exercise or one which might be pursued along the path of identifying, in the case of an
- 10 orphan receptor, an endogenous ligand. Tissue scans can be conducted across a broad range of healthy and diseased tissues. Such tissue scans provide a preferred first step in associating a specific receptor with a disease and/or disorder.

Preferably, the DNA sequence of the endogenous GPCR is used to make a probe for either radiolabeled cDNA or RT-PCR identification of the expression of the GPCR in tissue samples.

- 15 The presence of a receptor in a diseased tissue, or the presence of the receptor at elevated or decreased concentrations in diseased tissue compared to a normal tissue, can be preferably utilized to identify a correlation with that disease. Receptors can equally well be localized to regions of organs by this technique. Based on the known functions of the specific tissues to which the receptor is localized, the putative functional role of the receptor can be deduced.

20 **C. A "Human GPCR Proline Marker" Algorithm and the Creation of Non-Endogenous, Constitutively-Active Human GPCRs**

Among the many challenges facing the biotechnology arts is the unpredictability in gleaning genetic information from one species and correlating that information to another species

- 21 -

- nowhere in this art does this problem evidence more annoying exacerbation than in the genetic sequences that encode nucleic acids and proteins. Thus, for consistency and because of the highly unpredictable nature of this art, the following invention is limited, in terms of mammals, to human GPCRs - applicability of this invention to other mammalian species, while a potential possibility, is considered beyond mere rote application.

In general, when attempting to apply common "rules" from one related protein sequence to another or from one species to another, the art has typically resorted to sequence alignment, *i.e.*, sequences are linearized and attempts are then made to find regions of commonality between two or more sequences. While useful, this approach does not always prove to result in meaningful information. In the case of GPCRs, while the general structural motif is identical for all GPCRs, the variations in lengths of the TMs, ECs and ICs make such alignment approaches from one GPCR to another difficult at best. Thus, while it may be desirable to apply a consistent approach to, *e.g.*, constitutive activation from one GPCR to another, because of the great diversity in sequence length, fidelity, etc from one GPCR to the next, a generally applicable, and readily successful mutational alignment approach is in essence not possible. In an analogy, such an approach is akin to having a traveler start a journey at point A by giving the traveler dozens of different maps to point B, without any scale or distance markers on any of the maps, and then asking the traveler to find the shortest and most efficient route to destination B only by using the maps. In such a situation, the task can be readily simplified by having (a) a common "place-marker" on each map, and (b) the ability to measure the distance from the place-marker to destination B - this, then, will allow the traveler to select the most efficient from starting-point A to destination B.

In essence, a feature of the invention is to provide such coordinates within human GPCRs

that readily allows for creation of a constitutively active form of the human GPCRs.

As those in the art appreciate, the transmembrane region of a cell is highly hydrophobic; thus, using standard hydrophobicity plotting techniques, those in the art are readily able to determine the TM regions of a GPCR, and specifically TM6 (this same approach is also
5 applicable to determining the EC and IC regions of the GPCR). It has been discovered that within the TM6 region of human GPCRs, a common proline residue (generally near the middle of TM6), acts as a constitutive activation "marker." By counting 15 amino acids from the proline marker, the 16th amino acid (which is located in the IC3 loop), when mutated from its endogenous form to a non-endogenous form, leads to constitutive activation of the receptor. For convenience, we
10 refer to this as the "Human GPCR Proline Marker" Algorithm. Although the non-endogenous amino acid at this position can be any of the amino acids, most preferably, the non-endogenous amino acid is lysine. While not wishing to be bound by any theory, we believe that this position itself is unique and that the mutation at this location impacts the receptor to allow for constitutive activation.

15 We note that, for example, when the endogenous amino acid at the 16th position is already lysine (as is the case with GPR4 and GPR32), then in order for X to be a non-endogenous amino acid, it must be other than lysine; thus, in those situations where the endogenous GPCR has an endogenous lysine residue at the 16th position, the non-endogenous version of that GPCR preferably incorporates an amino acid other than lysine, preferably alanine, histidine and arginine,
20 at this position. Of further note, it has been determined that GPR4 appears to be linked to Gs and active in its endogenous form (data not shown).

Because there are only 20 naturally occurring amino acids (although the use of non-naturally occurring amino acids is also viable), selection of a particular non-endogenous amino

- 23 -

acid for substitution at this 16th position is viable and allows for efficient selection of a non-endogenous amino acid that fits the needs of the investigator. However, as noted, the more preferred non-endogenous amino acids at the 16th position are lysine, histidine, arginine and alanine, with lysine being most preferred. Those of ordinary skill in the art are credited with the ability to readily determine proficient methods for changing the sequence of a codon to achieve a desired mutation.

It has also been discovered that occasionally, but not always, the proline residue marker will be preceded in TM6 by W2 (*i.e.*, W2P¹AA₁₅X) where W is tryptophan and 2 is any amino acid residue.

Our discovery, amongst other things, negates the need for unpredictable and complicated sequence alignment approaches commonly used by the art. Indeed, the strength of our discovery, while an algorithm in nature, is that it can be applied in a facile manner to human GPCRs, with dexterous simplicity by those in the art, to achieve a unique and highly useful end-product, *i.e.*, a constitutively activated version of a human GPCR. Because many years and significant amounts of money will be required to determine the endogenous ligands for the human GPCRs that the Human Genome project is uncovering, the disclosed invention not only reduces the time necessary to positively exploit this sequence information, but at significant cost-savings. This approach truly validates the importance of the Human Genome Project because it allows for the utilization of genetic information to not only understand the role of the GPCRs in, *e.g.*, diseases, but also provides the opportunity to improve the human condition.

D. Screening of Candidate Compounds

1. Generic GPCR screening assay techniques

When a G protein receptor becomes constitutively active, it couples to a G protein (*e.g.*,

Gq, Gs, Gi, Go) and stimulates release and subsequent binding of GTP to the G protein. The G protein then acts as a GTPase and slowly hydrolyzes the GTP to GDP, whereby the receptor, under normal conditions, becomes deactivated. However, constitutively activated receptors, including the non-endogenous, human constitutively active GPCRs of the present invention, continue to exchange GDP for GTP. A non-hydrolyzable analog of GTP, [³⁵S]GTPγS, can be used to monitor enhanced binding to G proteins present on membranes which express constitutively activated receptors. It is reported that [³⁵S]GTPγS can be used to monitor G protein coupling to membranes in the absence and presence of ligand. An example of this monitoring, among other examples well-known and available to those in the art, was reported by Traynor and Nahorski in 1995. The preferred use of this assay system is for initial screening of candidate compounds because the system is generically applicable to all G protein-coupled receptors regardless of the particular G protein that interacts with the intracellular domain of the receptor.

B 2. Specific GPCR screening assay techniques

C Once candidate compounds are identified using the "generic" G protein-coupled receptor assay (*i.e.*, an assay to select compounds that are agonists, partial agonists, or inverse agonists), further screening to confirm that the compounds have interacted at the receptor site is preferred. For example, a compound identified by the "generic" assay may not bind to the receptor, but may instead merely "uncouple" the G protein from the intracellular domain.

a. Gs and Gi.

Gs stimulates the enzyme adenylyl cyclase. Gi (and Go), on the other hand, inhibit this enzyme. Adenylyl cyclase catalyzes the conversion of ATP to cAMP; thus,

- 25 -

constitutively activated GPCRs that couple the Gs protein are associated with increased cellular levels of cAMP. On the other hand, constitutively activated GPCRs that couple the Gi (or Go) protein are associated with decreased cellular levels of cAMP. *See, generally, "Indirect Mechanisms of Synaptic Transmission," Chpt. 8, From Neuron To Brain (3rd Ed.)*

5 Nichols, J.G. et al eds. Sinauer Associates, Inc. (1992). Thus, assays that detect cAMP can be utilized to determine if a candidate compound is, *e.g.*, an inverse agonist to the receptor (*i.e.*, such a compound would decrease the levels of cAMP). A variety of approaches known in the art for measuring cAMP can be utilized; a most preferred approach relies upon the use of anti-cAMP antibodies in an ELISA-based format. Another type of assay that can be
10 utilized is a whole cell second messenger reporter system assay. Promoters on genes drive the expression of the proteins that a particular gene encodes. Cyclic AMP drives gene expression by promoting the binding of a cAMP-responsive DNA binding protein or transcription factor (CREB) which then binds to the promoter at specific sites called cAMP response elements and drives the expression of the gene. Reporter systems can be constructed which have a promoter containing
15 multiple cAMP response elements before the reporter gene, *e.g.*, β -galactosidase or luciferase. Thus, a constitutively activated Gs-linked receptor causes the accumulation of cAMP that then activates the gene and expression of the reporter protein. The reporter protein such as β -galactosidase or luciferase can then be detected using standard biochemical assays (Chen et al. 1995). With respect to GPCRs that link to Gi (or Go), and thus decrease levels of cAMP, an
20 approach to the screening of, *e.g.*, inverse agonists, based upon utilization of receptors that link to Gs (and thus increase levels of cAMP) is disclosed in the Example section with respect to GPR17 and GPR30.

b. Go and Gq.

Gq and Go are associated with activation of the enzyme phospholipase C, which in turn hydrolyzes the phospholipid PIP_2 , releasing two intracellular messengers: diacylglycerol (DAG) and inistol 1,4,5-triphoisphate (IP_3). Increased accumulation of IP_3 is associated with activation of Gq- and Go-associated receptors. *See, generally*, "Indirect Mechanisms of Synaptic Transmission," Chpt. 8, From Neuron To Brain (3rd Ed.) Nichols, J.G. et al eds. Sinauer Associates, Inc. (1992). Assays that detect IP_3 accumulation can be utilized to determine if a candidate compound is, *e.g.*, an inverse agonist to a Gq- or Go-associated receptor (*i.e.*, such a compound would decrease the levels of IP_3). Gq-associated receptors can also been examined using an AP1 reporter assay in that Gq-dependent phospholipase C causes activation of genes containing AP1 elements; thus, activated Gq-associated receptors will evidence an increase in the expression of such genes, whereby inverse agonists thereto will evidence a decrease in such expression, and agonists will evidence an increase in such expression. Commercially available assays for such detection are available.

E. Medicinal Chemistry

Generally, but not always, direct identification of candidate compounds is preferably conducted in conjunction with compounds generated via combinatorial chemistry techniques, whereby thousands of compounds are randomly prepared for such analysis. Generally, the results of such screening will be compounds having unique core structures; thereafter, these compounds are preferably subjected to additional chemical modification around a preferred core structure(s) to further enhance the medicinal properties thereof. Such techniques are

- 27 -

known to those in the art and will not be addressed in detail in this patent document.

F. Pharmaceutical Compositions

Candidate compounds selected for further development can be formulated into pharmaceutical compositions using techniques well known to those in the art. Suitable
5 pharmaceutically-acceptable carriers are available to those in the art; for example, *see* Remington's Pharmaceutical Sciences, 16th Edition, 1980, Mack Publishing Co., (Oslo et al., eds.)

G. Other Utility

Although a preferred use of the non-endogenous versions of the disclosed human GPCRs is for the direct identification of candidate compounds as inverse agonists, agonists or partial
10 agonists (preferably for use as pharmaceutical agents), these receptors can also be utilized in research settings. For example, in vitro and in vivo systems incorporating these receptors can be utilized to further elucidate and understand the roles of the receptors in the human condition, both normal and diseased, as well understanding the role of constitutive activation as it applies to understanding the signaling cascade. A value in these non-endogenous receptors is that their
15 utility as a research tool is enhanced in that, because of their unique features, the disclosed receptors can be used to understand the role of a particular receptor in the human body before the endogenous ligand therefor is identified. Other uses of the disclosed receptors will become apparent to those in the art based upon, *inter alia*, a review of this patent document.

EXAMPLES

20 The following examples are presented for purposes of elucidation, and not limitation, of the present invention. Following the teaching of this patent document that a mutational cassette may be utilized in the IC3 loop of human GPCRs based upon a position relative to a proline residue in TM6 to constitutively activate the receptor, and while specific nucleic acid

and amino acid sequences are disclosed herein, those of ordinary skill in the art are credited with the ability to make minor modifications to these sequences while achieving the same or substantially similar results reported below. Particular approaches to sequence mutations are within the purview of the artisan based upon the particular needs of the artisan.

5 **Example 1**

Preparation of Endogenous Human GPCRs

A variety of GPCRs were utilized in the Examples to follow. Some endogenous human GPCRs were graciously provided in expression vectors (as acknowledged below) and other endogenous human GPCRs were synthesized *de novo* using publicly-available sequence
10 information.

1. **GPR1 (GenBank Accession Number: U13666)**

The human cDNA sequence for GPR1 was provided in pRcCMV by Brian O'Dowd (University of Toronto). GPR1 cDNA (1.4kB fragment) was excised from the pRcCMV vector as a NdeI-XbaI fragment and was subcloned into the NdeI-XbaI site of pCMV vector (*see*
15 Figure 3). Nucleic acid (SEQ.ID.NO.: 1) and amino acid (SEQ.ID.NO.: 2) sequences for human GPR1 were thereafter determined and verified.

2. **GPR4 (GenBank Accession Numbers: L36148, U35399, U21051)**

The human cDNA sequence for GPR4 was provided in pRcCMV by Brian O'Dowd (University of Toronto). GPR1 cDNA (1.4kB fragment) was excised from the pRcCMV
20 vector as an ApaI(blunted)-XbaI fragment and was subcloned (with most of the 5' untranslated region removed) into HindIII(blunted)-XbaI site of pCMV vector. Nucleic acid (SEQ.ID.NO.: 3) and amino acid (SEQ.ID.NO.: 4) sequences for human GPR4 were thereafter determined and verified.

3. **GPR5 (GenBank Accession Number: L36149)**

- 29 -

The cDNA for human GPR5 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 64°C for 1min; and 72°C for 1.5 min. The 5' PCR primer contained an EcoRI site with the sequence: 5'-TATGAATTCAGATGCTCTAAACGTCCCTGC-3' (SEQ.ID.NO.: 5) and the 3' primer contained BamHI site with the sequence: 5'-TCCGGATCCACCTGCACCTGCGCCTGCACC-3' (SEQ.ID.NO.: 6). The 1.1 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 7) and amino acid (SEQ.ID.NO.: 8) sequences for human GPR5 were thereafter determined and verified.

4. GPR7 (GenBank Accession Number: U22491)

The cDNA for human GPR7 was generated and cloned into pCMV expression vector as follows: PCR condition- PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 62°C for 1min; and 72°C for 1min and 20 sec. The 5' PCR primer contained a HindIII site with the sequence: 5'-GCAAGCTTGGGGGACGCCAGGTCGCCGGCT-3' (SEQ.ID.NO.: 9) and the 3' primer contained a BamHI site with the sequence: 5'-GCGGATCCGGACGCTGGGGGAGTCAGGCTGC-3' (SEQ.ID.NO.: 10). The 1.1 kb PCR fragment was digested with HindIII and BamHI and cloned into HindIII-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 11) and amino acid (SEQ.ID.NO.:

12) sequences for human GPR7 were thereafter determined and verified.

5. GPR8 (GenBank Accession Number: U22492)

The cDNA for human GPR8 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 62°C for 1min; and 72 °C for 1min and 20 sec. The 5' PCR primer contained an EcoRI site with the sequence:

5'-CGGAATTCGTCAACGGTCCCAGCTACAATG-3' (SEQ.ID.NO.: 13).

10 and the 3' primer contained a BamHI site with the sequence:

5'-ATGGATCCCAGGCCCTTCAGCACCGCAATAT-3'(SEQ.ID.NO.: 14).

The 1.1 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of pCMV expression vector. All 4 cDNA clones sequenced contained a possible polymorphism involving a change of amino acid 206 from Arg to Gln. Aside from this

15 difference, nucleic acid (SEQ.ID.NO.: 15) and amino acid (SEQ.ID.NO.: 16) sequences for human GPR8 were thereafter determined and verified.

6. GPR9 (GenBank Accession Number: X95876)

The cDNA for human GPR9 was generated and cloned into pCMV expression vector as follows: PCR was performed using a clone (provided by Brian O'Dowd) as template and pfu polymerase (Stratagene) with the buffer system provided by the manufacturer supplemented with 10% DMSO, 0.25 μ M of each primer, and 0.5 mM of each of the 4 nucleotides. The cycle condition was 25 cycles of: 94°C for 1 min; 56°C for 1min; and 72 °C for 2.5 min. The 5' PCR primer contained an EcoRI site with the sequence:

- 31 -

5'-ACGAATTCAGCCATGGTCCTTGAGGTGAGTGACCACCAAGTGCTAAAT-3'
(SEQ.ID.NO.: 17)

and the 3' primer contained a BamHI site with the sequence:

5'-GAGGATCCTGGAATGCGGGGAAGTCAG-3' (SEQ.ID.NO.: 18).

- 5 The 1.2 kb PCR fragment was digested with EcoRI and cloned into EcoRI-SmaI site of PCMV expression vector. Nucleic acid (SEQ.ID.NO.: 19) and amino acid (SEQ.ID.NO.: 20) sequences for human GPR9 were thereafter determined and verified.

7. GPR9-6 (GenBank Accession Number: U45982)

The cDNA for human GPR9-6 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 62°C for 1min; and 72°C for 1 min and 20 sec. The 5' PCR primer was kinased with the sequence:
5'-TTAAGCTTGACCTAATGCCATCTTGTGTCC-3' (SEQ.ID.NO.: 21)

- 15 and the 3' primer contained a BamHI site with the sequence:

5'-TTGGATCCAAAAGAACCATGCACCTCAGAG-3' (SEQ.ID.NO.: 22).

The 1.2 kb PCR fragment was digested with BamHI and cloned into EcoRV-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 23) and amino acid (SEQ.ID.NO.: 24) sequences for human GPR9-6 were thereafter determined and verified.

20 8. GPR10 (GenBank Accession Number: U32672)

The human cDNA sequence for GPR10 was provided in pRcCMV by Brian O'Dowd (University of Toronto). GPR10 cDNA (1.3kB fragment) was excised from the pRcCMV vector as an EcoRI-XbaI fragment and was subcloned into EcoRI-XbaI site of pCMV

vector. Nucleic acid (SEQ.ID.NO.: 25) and amino acid (SEQ.ID.NO.: 26) sequences for human GPR10 were thereafter determined and verified.

9. GPR15 (GenBank Accession Number: U34806)

The human cDNA sequence for GPR15 was provided in pCDNA3 by Brian O'Dowd (University of Toronto). GPR15 cDNA (1.5kB fragment) was excised from the pCDNA3 vector as a HindIII-Bam fragment and was subcloned into HindIII-Bam site of pCMV vector. Nucleic acid (SEQ.ID.NO.: 27) and amino acid (SEQ.ID.NO.: 28) sequences for human GPR15 were thereafter determined and verified.

10. GPR17 (GenBank Accession Number: Z94154)

10 The cDNA for human GPR17 was generated and cloned into pCMV expression
vector as follows: PCR was performed using genomic DNA as template and rTth polymerase
(Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and
0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 56°C for
1min and 72 °C for 1 min and 20 sec. The 5' PCR primer contained an EcoRI site with the
15 sequence:

5'-CTAGAATTCTGACTCCAGCCAAAGCATGAAT-3' (SEQ.ID.NO.: 29) and the 3' primer contained a BamHI site with the sequence:

5'-GCTGGATCCTAAACAGTCTGCGCTCGGCCT-3' (SEQ.ID.NO.: 30).

The 1.1 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI
20 site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 31) and amino acid (SEQ.ID.NO.:
32) sequences for human GPR17 were thereafter determined and verified.

11. GPR18 (GenBank Accession Number: L42324)

The cDNA for human GPR18 was generated and cloned into pCMV expression

- 33 -

vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 54°C for 1min; and 72°C for 1min and 20 sec. The 5' PCR primer was kinased with the sequence:

5 5'-ATAAGATGATCACCCTGAACAATCAAGAT -3' (SEQ.ID.NO.: 33)

and the 3' primer contained an EcoRI site with the sequence:

5'-TCCGAATTCATAACATTTCACTGTTTATATTGC-3' (SEQ.ID.NO.: 34).

The 1.0 kb PCR fragment was digested with EcoRI and cloned into blunt-EcoRI site of pCMV expression vector. All 8 cDNA clones sequenced contained 4 possible polymorphisms involving:
10 changes of amino acid 12 from Thr to Pro, amino acid 86 from Ala to Glu, amino acid 97 from Ile to Leu and amino acid 310 from Leu to Met. Aside from these changes, nucleic acid (SEQ.ID.NO.: 35) and amino acid (SEQ.ID.NO.: 36) sequences for human GPR18 were thereafter determined and verified.

12. GPR20 (GenBank Accession Number: U66579)

15 The cDNA for human GPR20 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 62°C for 1min; and 72°C for 1 min and 20 sec. The 5' PCR primer was kinased with the sequence:

20 5'-CCAAGCTTCCAGGCCTGGGGTGTGCTGG-3' (SEQ.ID.NO.: 37)

and the 3' primer contained a BamHI site with the sequence:

5'-ATGGATCCTGACCTTCGGCCCCTGGCAGA-3' (SEQ.ID.NO.: 38).

The 1.2 kb PCR fragment was digested with BamHI and cloned into EcoRV-BamHI site of

PCMV expression vector. Nucleic acid (SEQ.ID.NO.: 39) and amino acid (SEQ.ID.NO.: 40) sequences for human GPR20 were thereafter determined and verified.

13. GPR21 (GenBank Accession Number: U66580)

The cDNA for human GPR21 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 62°C for 1min; and 72 °C for 1 min and 20 sec. The 5' PCR primer was kinased with the sequence:

5'-GAGAATTCACTCCTGAGCTCAAGATGAACT-3' (SEQ.ID.NO.: 41)

and the 3' primer contained a BamHI site with the sequence:

5'-CGGGATCCCCGTAAGTCTGAGCCACTTCAGAT-3' (SEQ.ID.NO.: 42).

The 1.1 kb PCR fragment was digested with BamHI and cloned into EcoRV-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 43) and amino acid (SEQ.ID.NO.: 44) sequences for human GPR21 were thereafter determined and verified.

14. GPR22 (GenBank Accession Number: U66581)

The cDNA for human GPR22 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 50°C for 1min; and 72 °C for 1.5 min. The 5' PCR primer was kinased with the sequence:

5'-TCCCCCGGGAAAAAACC AACTGCTCCAAA-3' (SEQ.ID.NO.: 45)

and the 3' primer contained a BamHI site with the sequence:

5'-TAGGATCCATTTGAATGTGGATTTGGTAAA-3' (SEQ.ID.NO.: 46).

- 35 -

The 1.38 kb PCR fragment was digested with BamHI and cloned into EcoRV-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 47) and amino acid (SEQ.ID.NO.: 48) sequences for human GPR22 were thereafter determined and verified.

15. GPR24 (GenBank Accession Number: U71092)

5 The cDNA for human GPR24 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 56°C for 1min; and 72 °C for 1 min and 20 sec. The 5' PCR primer contains a HindIII site with the
10 sequence:

5'-GTGAAGCTTGCCTCTGGTGCCTGCAGGAGG-3' (SEQ.ID.NO.: 49)

and the 3' primer contains an EcoRI site with the sequence:

5'-GCAGAATTCCTGGTGGCGTGTGTGGTGCCC-3' (SEQ.ID.NO.: 50).

The 1.3 kb PCR fragment was digested with HindIII and EcoRI and cloned into HindIII-EcoRI
15 site of pCMV expression vector. The nucleic acid (SEQ.ID.NO.: 51) and amino acid sequence (SEQ.ID.NO.: 52) for human GPR24 were thereafter determined and verified.

16. GPR30 (GenBank Accession Number: U63917)

The cDNA for human GPR30 was generated and cloned as follows: the coding sequence of GPR30 (1128bp in length) was amplified from genomic DNA using the primers:

20 5'-GGCGGATCCATGGATGTGACTTCCCAA-3' (SEQ.ID.NO.: 53) and

5'-GGCGGATCCCTACACGGCACTGCTGAA-3' (SEQ.ID.NO.: 54).

The amplified product was then cloned into a commercially available vector, pCR2.1 (Invitrogen),

using a "TOPO-TA Cloning Kit" (Invitrogen, #K4500-01), following manufacturer instructions. The full-length GPR30 insert was liberated by digestion with BamHI, separated from the vector by agarose gel electrophoresis, and purified using a Sephaglas Bandprep™ Kit (Pharmacia, # 27-9285-01) following manufacturer instructions. The nucleic acid (SEQ.ID.NO.: 55) and amino acid
5 sequence (SEQ.ID.NO.: 56) for human GPR30 were thereafter determined and verified.

17. GPR31 (GenBank Accession Number: U65402)

The cDNA for human GPR31 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and
10 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 58°C for 1min; and 72 °C for 2 min. The 5' PCR primer contained an EcoRI site with the sequence:
5'-AAGGAATTCACGGCCGGGTGATGCCATTCCC-3' (SEQ.ID.NO.: 57)
and the 3' primer contained a BamHI site with the sequence:
5'-GGTGGATCCATAAACACGGGCGTTGAGGAC -3' (SEQ.ID.NO.: 58).

15 The 1.0 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 59) and amino acid (SEQ.ID.NO.: 60) sequences for human GPR31 were thereafter determined and verified.

18. GPR32 (GenBank Accession Number: AF045764)

The cDNA for human GPR32 was generated and cloned into pCMV expression
20 vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 56°C for 1min; and 72 °C for 1 min and 20 sec. The 5' PCR primer contained an EcoRI site with the

- 37 -

sequence:

5'-TAAGAATTCCATAAAAAATTATGGAATGG-3' (SEQ.ID.NO.:243)

and the 3' primer contained a BamHI site with the sequence:

5'-CCAGGATCCAGCTGAAGTCTTCCATCATTC-3' (SEQ.ID.NO.: 244).

- 5 The 1.1 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 245) and amino acid (SEQ.ID.NO.: 246) sequences for human GPR32 were thereafter determined and verified.

19. GPR40 (GenBank Accession Number: AF024687)

- 10 The cDNA for human GPR40 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min, 65°C for 1min and 72 °C for 1 min and 10 sec. The 5' PCR primer contained an EcoRI site with the sequence

- 15 5'-GCAGAATTCGGCGGCCCCCATGGACCTGCCCCC-3' (SEQ.ID.NO.: 247)

and the 3' primer contained a BamHI site with the sequence

5'-GCTGGATCCCCCGAGCAGTGGCGTTACTTC-3' (SEQ.ID.NO.: 248).

- The 1 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 249) and amino acid (SEQ.ID.NO.: 250) sequences for human GPR40 were thereafter determined and verified.
- 20

20. GPR41 (GenBank Accession Number AF024688)

The cDNA for human GPR41 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase

(Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of 94°C for 1 min, 65°C for 1min and 72 °C for 1 min and 10 sec. The 5' PCR primer contained an HindIII site with the sequence:

5 5'-CTCAAGCTTACTCTCTCTCACCAGTGGCCAC-3' (SEQ.ID.NO.: 251)

and the 3' primer was kinased with the sequence

5'-CCCTCCTCCCCCGGAGGACCTAGC-3' (SEQ.ID.NO.: 252).

The 1 kb PCR fragment was digested with HindIII and cloned into HindIII-blunt site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 253) and amino acid (SEQ.ID.NO.: 254)

10 sequences for human GPR41 were thereafter determined and verified.

21. GPR43 (GenBank Accession Number AF024690)

The cDNA for human GPR43 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and

15 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1min; and 72 °C for 1 min and 10 sec. The 5' PCR primer contains an HindIII site with the sequence:

5'-TTTAAGCTTCCCCTCCAGGATGCTGCCGGAC-3' (SEQ.ID.NO.: 255)

and the 3' primer contained an EcoRI site with the sequence:

20 5'-GGCGAATTCTGAAGGTCCAGGGAACTGCTA-3' (SEQ.ID.NO. 256).

The 1 kb PCR fragment was digested with HindIII and EcoRI and cloned into HindIII-EcoRI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 257) and amino acid (SEQ.ID.NO.: 258)

sequences for human GPR43 were thereafter determined and verified.

- 39 -

22. APJ (GenBank Accession Number: U03642)

Human APJ cDNA (in pRcCMV vector) was provided by Brian O'Dowd (University of Toronto). The human APJ cDNA was excised from the pRcCMV vector as an EcoRI-XbaI (blunted) fragment and was subcloned into EcoRI-SmaI site of pCMV vector.

- 5 Nucleic acid (SEQ.ID.NO.: 61) and amino acid (SEQ.ID.NO.: 62) sequences for human APJ were thereafter determined and verified.

23. BLR1 (GenBank Accession Number: X68149)

- The cDNA for human BLR1 was generated and cloned into pCMV expression vector as follows: PCR was performed using thymus cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 62°C for 1min; and 72 °C for 1 min and 20 sec. The 5' PCR primer contained an EcoRI site with the sequence:

5'-TGAGAATTCTGGTGACTCACAGCCGGCACAG-3' (SEQ.ID.NO.: 63);

- 15 and the 3' primer contained a BamHI site with the sequence:

5'-GCCGGATCCAAGGAAAAGCAGCAATAAAAGG-3' (SEQ.ID.NO.: 64). The 1.2 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 65) and amino acid (SEQ.ID.NO.: 66) sequences for human BLR1 were thereafter determined and verified.

20 24. CEPR (GenBank Accession Number: U77827)

The cDNA for human CEPR was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and

0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1min; and 72 °C for 1 min and 20 sec. The 5' PCR primer was kinased with the sequence:

5'-CAAAGCTTGAAAGCTGCACGGTGCAGAGAC-3' (SEQ.ID.NO.:67)

and the 3' primer contained a BamHI site with the sequence:

5'-GCGGATCCCGAGTCACACCCTGGCTGGGCC-3' (SEQ.ID.NO.: 68).

The 1.2 kb PCR fragment was digested with BamHI and cloned into EcoRV-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 69) and amino acid (SEQ.ID.NO.: 70) sequences for human CEPR were thereafter determined and verified.

25. EBI1 (GenBank Accession Number: L31581)

The cDNA for human EBI1 was generated and cloned into pCMV expression vector as follows: PCR was performed using thymus cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 62°C for 1min; and 72 °C for 1 min and 20 sec. The 5' PCR primer contained an EcoRI site with the sequence:

5'-ACAGAATTCCTGTGTGGTTTTACCGCCCAG-3' (SEQ.ID.NO.: 71)

and the 3' primer contained a BamHI site with the sequence:

5'-CTCGGATCCAGGCAGAAGAGTCGCCTATGG-3' (SEQ.ID.NO.: 72).

The 1.2 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 73) and amino acid (SEQ.ID.NO.: 74) sequences for human EBI1 were thereafter determined and verified.

26. EBI2 (GenBank Accession Number: L08177)

The cDNA for human EBI2 was generated and cloned into pCMV expression

- 41 -

vector as follows: PCR was performed using cDNA clone (graciously provided by Kevin Lynch, University of Virginia Health Sciences Center; the vector utilized was not identified by the source) as template and pfu polymerase (Stratagene) with the buffer system provided by the manufacturer supplemented with 10% DMSO, 0.25 μ M of each primer, and 0.5 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 60°C for 1min; and 72 °C for 1 min and 20 sec. The 5' PCR primer contained an EcoRI site with the sequence:

5'-CTGGAATTCACCTGGACCACCACCAATGGATA-3' (SEQ.ID.NO.: 75)

and the 3' primer contained a BamHI site with the sequence

5'-CTCGGATCCTGCAAAGTTTGTTCATACAG TT-3' (SEQ.ID.NO.: 76).

10 The 1.2 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 77) and amino acid (SEQ.ID.NO.: 78) sequences for human EBI2 were thereafter determined and verified.

27. ETBR-LP2 (GenBank Accession Number: D38449)

The cDNA for human ETBR-LP2 was generated and cloned into pCMV expression vector as follows: PCR was performed using brain cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1min; and 72 °C for 1.5 min. The 5' PCR contained an EcoRI site with the sequence:

20 5'-CTGGAATTCTCCTGCTCATCCAGCCATGCGG -3' (SEQ.ID.NO.: 79)

and the 3' primer contained a BamHI site with the sequence:

5'-CCTGGATCCCCACCCCTACTGGGGCCTCAG -3' (SEQ.ID.NO.: 80).

The 1.5 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI

site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 81) and amino acid (SEQ.ID.NO.: 82) sequences for human ETBR-LP2 were thereafter determined and verified.

28. GHSR (GenBank Accession Number: U60179)

The cDNA for human GHSR was generated and cloned into pCMV expression vector as follows: PCR was performed using hippocampus cDNA as template and TaqPlus Precision polymerase (Stratagene) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 68°C for 1min; and 72°C for 1 min and 10 sec. For first round PCR, the 5' PCR primer sequence was:

10 5'-ATGTGGAACGCGACGCCAGCG-3' (SEQ.ID.NO.: 83)

and the 3' primer sequence was:

5'-TCATGTATTAATACTAGATTCT-3' (SEQ.ID.NO.: 84).

Two microliters of the first round PCR was used as template for the second round PCR where the 5' primer was kinased with sequence:

15 5'-TACCATGTGGAACGCGACGCCAGCGAAGAGCCGGGGT-3'(SEQ.ID.NO.:85)

and the 3' primer contained an EcoRI site with the sequence:

5'-CGGAATTCATGTATTAATACTAGATTCTGTCCAGGCCCG-3'(SEQ.ID.NO.:86).

The 1.1 kb PCR fragment was digested with EcoRI and cloned into blunt-EcoRI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 87) and amino acid (SEQ.ID.NO.: 88) sequences
20 for human GHSR were thereafter determined and verified.

29. GPCR-CNS (GenBank Accession Number: AF017262)

The cDNA for human GPCR-CNS was generated and cloned into pCMV expression vector as follows: PCR was performed using brain cDNA as template and rTth

- 43 -

polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1min; and 72°C for 2 min. The 5' PCR primer contained a HindIII site with the sequence:

5 5'-GCAAGCTTGTGCCCTCACCAAGCCATGCGAGCC-3' (SEQ.ID.NO.: 89)

and the 3' primer contained an EcoRI site with the sequence:

5'-CGGAATTCAGCAATGAGTTCCGACAGAAGC-3' (SEQ.ID.NO.: 90).

The 1.9 kb PCR fragment was digested with HindIII and EcoRI and cloned into HindIII-EcoRI site of pCMV expression vector. All nine clones sequenced contained a potential polymorphism
10 involving a S284C change. Aside from this difference, nucleic acid (SEQ.ID.NO.: 91) and amino acid (SEQ.ID.NO.: 92) sequences for human GPCR-CNS were thereafter determined and verified.

30. GPR-NGA (GenBank Accession Number: U55312)

The cDNA for human GPR-NGA was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth
15 polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of 94°C for 1 min, 56°C for 1min and 72°C for 1.5 min. The 5' PCR primer contained an EcoRI site with the sequence:

5'-CAGAATTCAGAGAAAAAAGTGAATATGGTTTTT-3' (SEQ.ID.NO.: 93)

20 and the 3' primer contained a BamHI site with the sequence:

5'-TTGGATCCCTGGTGCATAACAATTGAAAGAAT-3' (SEQ.ID.NO.: 94).

The 1.3 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 95) and amino acid (SEQ.ID.NO.:

96) sequences for human GPR-NGA were thereafter determined and verified.

31. H9 (GenBank Accession Number: U52219)

The cDNA for human HB954 was generated and cloned into pCMV expression vector as follows: PCR was performed using pituitary cDNA as template and rTth polymerase

5 (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min, 62°C for 1min and 72°C for 2 min. The 5' PCR primer contains a HindIII site with the sequence:

5'-GGAAAGCTTAACGATCCCCAGGAGCAACAT-3' (SEQ.ID.NO.: 97)

and the 3' primer contains a BamHI site with the sequence:

10 5'-CTGGGATCCTACGAGAGCATTTTTCACACAG-3' (SEQ.ID.NO.: 98).

The 1.9 kb PCR fragment was digested with HindIII and BamHI and cloned into HindIII-BamHI site of pCMV expression vector. When compared to the published sequences, a different isoform with 12 bp in frame insertion in the cytoplasmic tail was also identified and designated "H9b." Both isoforms contain two potential polymorphisms involving changes

15 of amino acid P320S and amino acid G448A. Isoform H9a contained another potential polymorphism of amino acid S493N, while isoform H9b contained two additional potential polymorphisms involving changes of amino acid I502T and amino acid A532T (corresponding to amino acid 528 of isoform H9a). Nucleic acid (SEQ.ID.NO.: 99) and amino acid (SEQ.ID.NO.: 100) sequences for human H9 were thereafter determined and
20 verified (in the section below, both isoforms were mutated in accordance with the Human GPCR Proline Marker Algorithm).

32. HB954 (GenBank Accession Number: D38449)

The cDNA for human HB954 was generated and cloned into pCMV expression

- 45 -

vector as follows: PCR was performed using brain cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of 94°C for 1 min, 58°C for 1min and 72°C for 2 min. The 5' PCR contained a HindIII site with the sequence:

5 5'-TCCAAGCTTCGCCATGGGACATAACGGGAGCT -3' (SEQ.ID.NO.: 101)

and the 3' primer contained an EcoRI site with the sequence:

5'-CGTGAATTCCAAGAATTTACAATCCTTGCT -3' (SEQ.ID.NO.: 102).

The 1.6 kb PCR fragment was digested with HindIII and EcoRI and cloned into HindIII-EcoRI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 103) and amino acid
10 (SEQ.ID.NO.: 104) sequences for human HB954 were thereafter determined and verified.

33. HG38 (GenBank Accession Number: AF062006)

The cDNA for human HG38 was generated and cloned into pCMV expression
vector as follows: PCR was performed using brain cDNA as template and rTth polymerase (Perkin
Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM
15 of each 4 nucleotides. The cycle condition was 30 cycles of 94°C for 1 min, 56°C for 1min and
72 °C for 1 min and 30 sec. Two PCR reactions were performed to separately obtain the 5' and
3' fragment. For the 5' fragment, the 5' PCR primer contained an HindIII site with the sequence:

5'-CCCAAGCTTCGGGCACCATGGACACCTCCC-3' (SEQ.ID.NO.: 259)

and the 3' primer contained a BamHIsite with the sequence:

20 5'-ACAGGATCCAAATGCACAGCACTGGTAAGC-3' (SEQ.ID.NO.: 260).

This 5' 1.5 kb PCR fragment was digested with HindIII and BamHI and cloned into an HindIII-BamHI site of pCMV. For the 3' fragment, the 5' PCR primer was kinased with the sequence:

5'-CTATAACTGGGTTACATGGTTTAAC-3' (SEQ.ID.NO. 261)

and the 3' primer contained an EcoRI site with the sequence:

5'-TTTGAATTCACATATTAATTAGAGACATGG-3' (SEQ.ID.NO.: 262).

The 1.4 kb 3' PCR fragment was digested with EcoRI and subcloned into a blunt-EcoRI site of pCMV vector. The 5' and 3' fragments were then ligated together through a common EcoRV site
5 to generate the full length cDNA clone. Nucleic acid (SEQ.ID.NO.: 263) and amino acid (SEQ.ID.NO.: 264) sequences for human HG38 were thereafter determined and verified.

34. HM74 (GenBank Accession Number: D10923)

The cDNA for human HM74 was generated and cloned into pCMV expression vector as follows: PCR was performed using either genomic DNA or thymus cDNA (pooled) as
10 template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1min; and 72°C for 1 min and 20 sec. The
5' PCR primer contained an EcoRI site with the sequence:

5'-GGAGAATTCAGTAGGCGAGGCGCTCCATC-3' (SEQ.ID.NO.: 105)

15 and the 3' primer was kinased with the sequence:

5'-GGAGGATCCAGGAAACCTTAGGCCGAGTCC-3' (SEQ.ID.NO.:106).

The 1.3 kb PCR fragment was digested with EcoRI and cloned into EcoRI-SmaI site of pCMV expression vector. Clones sequenced revealed a potential polymorphism involving a N94K change. Aside from this difference, nucleic acid (SEQ.ID.NO.: 107) and amino acid
20 (SEQ.ID.NO.: 108) sequences for human HM74 were thereafter determined and verified.

35. MIG (GenBank Accession Numbers: AFO44600 and AFO44601)

The cDNA for human MIG was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and TaqPlus Precision

- 47 -

polymerase (Stratagene) for first round PCR or pfu polymerase (Stratagene) for second round PCR with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM (TaqPlus Precision) or 0.5 mM (pfu) of each of the 4 nucleotides. When pfu was used, 10% DMSO was included in the buffer. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C
5 for 1min; and 72 °C for: (a) 1 min for first round PCR; and (b) 2 min for second round PCR. Because there is an intron in the coding region, two sets of primers were separately used to generate overlapping 5' and 3' fragments. The 5' fragment PCR primers were:

5'-ACCATGGCTTGCAATGGCAGTGCGGCCAGGGGGCACT-3' (external sense)
(SEQ.ID.NO.: 109)

10 and

5'-CGACCAGGACAAACAGCATCTTGGTCACTTGTCTCCGGC-3' (internal antisense)
(SEQ.ID.NO.: 110).

The 3' fragment PCR primers were:

5'-GACCAAGATGCTGTTTGTCTCTGGTCGTGGTGTCTTGGCAT-3' (internal sense)
15 (SEQ.ID.NO.: 111) and

5'-CGGAATTCAGGATGGATCGGTCTCTTGCTGCGCCT-3' (external antisense with an
EcoRI site) (SEQ.ID.NO.: 112).

The 5' and 3' fragments were ligated together by using the first round PCR as template and the
kinased external sense primer and external antisense primer to perform second round PCR. The
20 1.2 kb PCR fragment was digested with EcoRI and cloned into the blunt-EcoRI site of pCMV
expression vector. Nucleic acid (SEQ.ID.NO.: 113) and amino acid (SEQ.ID.NO.: 114)
sequences for human MIG were thereafter determined and verified.

36. OGR1 (GenBank Accession Number: U48405)

The cDNA for human OGR1 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1min; and 72°C for 1 min and 20 sec. The 5' PCR primer was kinased with the sequence: 5'-GGAAGCTTCAGGCCCAAGATGGGGAACAT-3' (SEQ.ID.NO.: 115); and the 3' primer contained a BamHI site with the sequence: 5'-GTGGATCCACCCGCGGAGGACCCAGGCTAG -3' (SEQ.ID.NO.: 116). The 1.1 kb PCR fragment was digested with BamHI and cloned into the EcoRV-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 117) and amino acid (SEQ.ID.NO.: 118) sequences for human OGR1 were thereafter determined and verified.

37. Serotonin 5HT_{2A}

The cDNA encoding endogenous human 5HT_{2A} receptor was obtained by RT-PCR using human brain poly-A⁺ RNA; a 5' primer from the 5' untranslated region with an Xho I restriction site:

5'-GACCTCGAGTCCTTCTACACCTCATC-3' (SEQ.ID.NO: 119)

and a 3' primer from the 3' untranslated region containing an Xba I site:

5'-TGCTCTAGATTCCAGATAGGTGAAAACCTTG-3' (SEQ.ID.NO: 120)

PCR was performed using either TaqPlus™ precision polymerase (Stratagene) or rTth™ polymerase (Perkin Elmer) with the buffer system provided by the manufacturers, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 57°C for 1min; and 72°C for 2 min. The 1.5 kb PCR fragment was digested with Xba I and subcloned into Eco RV-Xba I site of pBluescript. The resulting cDNA clones were fully

- 49 -

sequenced and found to encode two amino acid changes from the published sequences. The first one was a T25N mutation in the N-terminal extracellular domain; the second is an H452Y mutation. Because cDNA clones derived from two independent PCR reactions using Taq polymerase from two different commercial sources (TaqPlus™ from Stratagene and rTth™ Perkin Elmer) contained the same two mutations, these mutations are likely to represent sequence polymorphisms rather than PCR errors. With these exceptions, the nucleic acid (SEQ.ID.NO.: 121) and amino acid (SEQ.ID.NO.: 122) sequences for human 5HT_{2A} were thereafter determined and verified.

38. Serotonin 5HT_{2C}

10 The cDNA encoding endogenous human 5HT_{2C} receptor was obtained from human brain poly-A⁺ RNA by RT-PCR. The 5' and 3' primers were derived from the 5' and 3' untranslated regions and contained the following sequences:

5'-GACCTCGAGGTTGCTTAAGACTGAAGC-3' (SEQ.ID.NO.: 123)

5'-ATTTCTAGACATATGTAGCTTGTACCG-3' (SEQ.ID.NO.: 124)

15 Nucleic acid (SEQ.ID.NO.: 125) and amino acid (SEQ.ID.NO.: 126) sequences for human 5HT_{2C} were thereafter determined and verified.

39. V28 (GenBank Accession Number: U20350)

20 The cDNA for human V28 was generated and cloned into pCMV expression vector as follows: PCR was performed using brain cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1 min; and 72°C for 1 min and 20 sec. The 5' PCR primer contained a HindIII site with the sequence: 5'-GGTAAGCTTGGCAGTCCACGCCAGGCCTTC-3' (SEQ.ID.NO.: 127)

and the 3' primer contained an EcoRI site with the sequence:

5'-TCCGAATTCTCTGTAGACACAAGGCTTTGG-3' (SEQ.ID.NO.: 128)

The 1.1 kb PCR fragment was digested with HindIII and EcoRI and cloned into HindIII-EcoRI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 129) and amino acid (SEQ.ID.NO.:

5 130) sequences for human V28 were thereafter determined and verified.

Example 2

PREPARATION OF NON-ENDOGENOUS HUMAN GPCRS

1. Site-Directed Mutagenesis

Mutagenesis based upon the Human GPCR Proline Marker approach disclosed herein was
10 performed on the foregoing endogenous human GPCRs using Transformer Site-Directed
Mutagenesis Kit (Clontech) according to the manufacturer instructions. For this mutagenesis
approach, a Mutation Probe and a Selection Marker Probe (unless otherwise indicated, the probe
of SEQ.ID.NO.: 132 was the same throughout) were utilized, and the sequences of these for the
specified sequences are listed below in Table B (the parenthetical number is the SEQ. ID.NO.).
15 For convenience, the codon mutation incorporated into the human GPCR is also noted, in standard
form:

Table B

Receptor Identifier (Codon Mutation)	Mutation Probe Sequence (5'-3') (SEQ.ID.NO.)	Selection Marker Probe Sequence (5'-3') (SEQ.ID.NO.)
20 GPR1 (F245K)	GATCTCCAGTAGGCATA <u>AAGT</u> GGACAATTCTGG (131)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAG (132)
GPR4 (K223A)	AGAAGGCCAAGATC <u>GCGCGG</u> CTGGCCCTCA (133)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
25 GPR5 (V224K)	CGGCGCCACCGCACG <u>AAAAA</u> GCTCATCTTC	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT

- 51 -

	(134)	
GPR7 (T250K)	GCCAAGAAGCGGGTGAAGTT CCTGGTGGTGGCA (135)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
GPR8 (T259K)	CAGGCGGAAGGTGAAAGTCC TGGTCCTCGT (136)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
30 GPR9 (M254K)	CGGCGCCTGCGGGCCAAAGCG GCTGGTGGTGGTG (137)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
GPR9-6 (L241K)	CCAAGCACAAAGCCAAGAAA GTGACCATCAC (138)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
GPR10 35 (F276K)	GCGCCGGCGCACCAAAATGCT TGCTGGTGGT (139)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
GPR15 (I240K)	CAAAAAGCTGAAGAAATCTA AGAAGATCATCTTTATTGTCG (140)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
GPR17 (V234K)	CAAGACCAAGGCAAAACGCA TGATCGCCAT (141)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
40 GPR18 (I231K)	GTCAAGGAGAAGTCCAAAAG GATCATCATC (142)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
GPR20 (M240K)	CGCCGCGTGCGGGCCAAAGCA GCTCCTGCTC (143)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
GPR21 45 (A251K)	CCTGATAAGCGCTATAAAAT GGTCCTGTTTCGA (144)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
GPR22 (F312K)	GAAAGACAAAAGAGAGTCA AGAGGATGTCTTTATTG (145)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
GPR24 (T304K)	CGGAGAAAGAGGGTGA AAC GCACAGCCATCGCC (146)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
50 GPR30 (L258K)	alternate approach; see below	alternate approach; see below
GPR31 (Q221K)	AAGCTTCAGCGGGCCAAAGGC ACTGGTCACC (147)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
GPR32 55 (K255A)	CATGCCAACCGGCCCCGCGAG GCTGCTGCTGGT (279)	ACCAGCAGCAGCCTCGCGGG CCGGTTGGCATG (280)
GPR40 (A223K)	CGGAAGCTGCGGGCCAAATG GGTGGCCGGC (265)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
GPR41	CAGAGGAGGGTGAAGGGGCT GTTGGCG	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT

(A223K)	(266)	
GPR43 (V221K)	GGCGGCGCCGAGCC <u>AAGGGG</u> CTGGCTGTGG (267)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
APJ (L247K)	alternate approach; see below	alternate approach; see below
BLR1 (V258K)	CAGCGGCAGAAGGCA <u>AAAA</u> GGGTGGCCATC (148)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
CEPR (L258K)	CGGCAGAAGGCGA <u>AAGCGCAT</u> GATCCTCGCG (149)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT

10 EBI1 (I262K)	GAGCGCAACAAGGCC <u>AAAA</u> AGGTGATCATC (150)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
EBI2 (L243K)	GGTGTAACAAAAAGGCTAA <u>AAACACAATTATTCTTATT</u> (151)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
15 ETBR-LP2 (N358K)	GAGAGCCAGCTCA <u>AAGAGCAC</u> CGTGGTG (152)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
GHSR (V262K)	CCACAAGCAAACCA <u>AAGAAAA</u> TGCTGGCTGT (153)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
GPCR-CNS (N491K)	CTAGAGAGTCAGATGA <u>AGTG</u> TACAGTAGTGGCAC (155)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
20 GPR-NGA (I275K)	CGGACAAAAGTGAAAACTAA <u>AAAGATGTTCTCTATT</u> (156)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
H9a and H9b (F236K)	GCTGAGGTTTCGAATAAACT AACCATGTTTGTG (157)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
25 HB954 (H265K)	GGGAGGCCGAGCTGA <u>AAAGCC</u> ACCCTGCTC (158)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
HG38 (V765K)	GGGACTGCTCTATGA <u>AAAAAA</u> CACATTGCCCTG (268)	CATCAAGTGTATCATGTGCC AAGTACGCC (154)

HM74 (I230K)	CAAGATCAAGAGAGCC <u>AAAA</u> CCTTCATCATG (159)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
30 MIG (T273K)	CCGGAGACAAGTGA <u>AGAAG</u> ATGCTGTTTGTG (160)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
OGR1 (Q227K)	GCAAGGACCAGATCA <u>AAGCGG</u> CTGGTGCTCA (161)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAAGT
35 Serotonin 5HT _{2A} (C322K)	alternate approach; see below	alternate approach; see below
Serotonin 5HT _{2C} (S310K)	alternate approach; see below	alternate approach; see below

- 53 -

V28 (I230K)	CAAGAAAGCCAAAGCCAAG AAACTGATCCTTCTG (162)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
----------------	---	-------------------------------------

The non-endogenous human GPCRs were then sequenced and the derived and verified nucleic acid and amino acid sequences are listed in the accompanying "Sequence Listing" appendix

5 to this patent document, as summarized in Table C below:

Table C

	Mutated GPCR	Nucleic Acid Sequence Listing	Amino Acid Sequence Listing
	GPR1 (F245K)	SEQ.ID.NO.: 163	SEQ.ID.NO.: 164
10	GPR4 (K223A)	SEQ.ID.NO.: 165	SEQ.ID.NO.: 166
	GPR5 (V224K)	SEQ.ID.NO.: 167	SEQ.ID.NO.: 168
	GPR7 (T250K)	SEQ.ID.NO.: 169	SEQ.ID.NO.: 170
15	GPR8 (T259K)	SEQ.ID.NO.: 171	SEQ.ID.NO.: 172
	GPR9 (M254K)	SEQ.ID.NO.: 173	SEQ.ID.NO.: 174
20	GPR9-6 (L241K)	SEQ.ID.NO.: 175	SEQ.ID.NO.: 176
	GPR10 (F276K)	SEQ.ID.NO.: 177	SEQ.ID.NO.: 178
	GPR15 (I240K)	SEQ.ID.NO.: 179	SEQ.ID.NO.: 180
25	GPR17 (V234K)	SEQ.ID.NO.: 181	SEQ.ID.NO.: 182
	GPR18 (I231K)	SEQ.ID.NO.: 183	SEQ.ID.NO.: 184
30	GPR20 (M240K)	SEQ.ID.NO.: 185	SEQ.ID.NO.: 186
	GPR21 (A251K)	SEQ.ID.NO.: 187	SEQ.ID.NO.: 188
	GPR22 (F312K)	SEQ.ID.NO.: 189	SEQ.ID.NO.: 190
35	GPR24 (T304K)	SEQ.ID.NO.: 191	SEQ.ID.NO.: 192
	GPR30	SEQ.ID.NO.: 193	SEQ.ID.NO.: 194

	(L258K)		
	GPR31 (Q221K)	SEQ.ID.NO.: 195	SEQ.ID.NO.: 196
5	GPR32 (K255A)	SEQ.ID.NO.: 269	SEQ.ID.NO.: 270
	GPR40 (A223K)	SEQ.ID.NO.: 271	SEQ.ID.NO.: 272
	GPR41 (A223K)	SEQ.ID.NO.: 273	SEQ.ID.NO.: 274
10	GPR43 (V221K)	SEQ.ID.NO.: 275	SEQ.ID.NO.: 276
	APJ (L247K)	SEQ.ID.NO.: 197	SEQ.ID.NO.: 198
15	BLR1 (V258K)	SEQ.ID.NO.: 199	SEQ.ID.NO.: 200
	CEPR (L258K)	SEQ.ID.NO.: 201	SEQ.ID.NO.: 202
	EBI1 (I262K)	SEQ.ID.NO.: 203	SEQ.ID.NO.: 204
20	EBI2 (L243K)	SEQ.ID.NO.: 205	SEQ.ID.NO.: 206
	ETBR-LP2 (N358K)	SEQ.ID.NO.: 207	SEQ.ID.NO.: 208
25	GHSR (V262K)	SEQ.ID.NO.: 209	SEQ.ID.NO.: 210
	GPCR-CNS (N491K)	SEQ.ID.NO.: 211	SEQ.ID.NO.: 212
	GPR-NGA (I275K)	SEQ.ID.NO.: 213	SEQ.ID.NO.: 214
30	H9a (F236K)	SEQ.ID.NO.: 215	SEQ.ID.NO.: 216
	H9b (F236K)	SEQ.ID.NO.: 217	SEQ.ID.NO.: 218
35	HB954 (H265K)	SEQ.ID.NO.: 219	SEQ.ID.NO.: 220
	HG38 (V765K)	SEQ.ID.NO.: 277	SEQ.ID.NO.: 278
	HM74 (I230K)	SEQ.ID.NO.: 221	SEQ.ID.NO.: 222
40	MIG (T273K)	SEQ.ID.NO.: 223	SEQ.ID.NO.: 224
	OGR1 (Q227K)	SEQ.ID.NO.: 225	SEQ.ID.NO.: 226
45	Serotonin 5HT _{2A} (C322K)	SEQ.ID.NO.: 227	SEQ.ID.NO.: 228
	Serotonin 5HT _{2C} (S310K)	SEQ.ID.NO.: 229	SEQ.ID.NO.: 230
	V28 (I230K)	SEQ.ID.NO.: 231	SEQ.ID.NO.: 232

- 55 -

2. Alternate Mutation Approaches for Employment of the Proline Marker Algorithm: APJ; Serotonin 5HT_{2A}; Serotonin 5HT_{2C}; and GPR30

Although the above site-directed mutagenesis approach is particularly preferred, other approaches can be utilized to create such mutations; those skilled in the art are readily credited with selecting approaches to mutating a GPCR that fits within the particular needs of the artisan.

a. APJ

Preparation of the non-endogenous, human APJ receptor was accomplished by mutating L247K. Two oligonucleotides containing this mutation were synthesized:

5'-GGCTTAAGAGCATCATCGTGGTGCTGGTG-3' (SEQ.ID.NO.: 233)

10 5'-GTCACCACCAGCACCACGATGATGCTCTTAAGCC-3' (SEQ.ID.NO.: 234)

The two oligonucleotides were annealed and used to replace the NaeI-BstEII fragment of human, endogenous APJ to generate the non-endogenous, version of human APJ.

b. Serotonin 5HT_{2A}

cDNA containing the point mutation C322K was constructed by utilizing the restriction enzyme site Sph I which encompasses amino acid 322. A primer containing the C322K mutation:

5'-CAAAGAAAGTACTGGGCATCGTCTTCTTCCT-3' (SEQ.ID.NO: 235)

was used along with the primer from the 3' untranslated region of the receptor:

5'-TGCTCTAGATTCCAGATAGGTGAAAA CTTG-3' (SEQ.ID.NO.: 236)

to perform PCR (under the conditions described above). The resulting PCR fragment was then used to replace the 3' end of endogenous 5HT_{2A} cDNA through the T4 polymerase blunted Sph I site.

c. Serotonin 5HT_{2C}

The cDNA containing a S310K mutation was constructed by replacing the Sty I restriction fragment containing amino acid 310 with synthetic double stranded oligonucleotides that encode

the desired mutation. The sense strand sequence utilized had the following sequence:

5'-CTAGGGGCACCATGCAGGCTATCAACAATGAAAGAAAAGCTAAGAAAGTC-3'

(SEQ. ID.NO.: 237)

and the antisense strand sequence utilized had the following sequence:

5 5'-CAAGGACTTTCTTAGCTTTTCTTTCATTGTTGATAGCCTGCATGGTGCCC-3' (SEQ.
ID. NO.: 238)

d. GPR30

Prior to generating non-endogenous GPR30, several independent pCR2.1/GPR30 isolates were sequenced in their entirety in order to identify clones with no PCR-generated mutations. A
10 clone having no mutations was digested with EcoRI and the endogenous GPR30 cDNA fragment was transferred into the CMV-driven expression plasmid pCI-neo (Promega), by digesting pCI-Neo with EcoRI and subcloning the EcoRI-liberated GPR30 fragment from pCR2.1/GPR30, to generate pCI/GPR30. Thereafter, the leucine at codon 258 was mutated to a lysine using a Quick-Change™ Site-Directed Mutagenesis Kit (Stratagene, #200518), according to manufacturer's
15 instructions, and the following primers:

5'-CGGCGGCAGAAGGCGAAACGCATGATCCTCGCGGT-3' (SEQ.ID.NO.: 239) and

5'-ACCGCGAGGATCATGCGTTTCGCCTTCTGC CGCCG-3' (SEQ.ID.NO.: 240)

Example 3

Receptor (Endogenous and Mutated) Expression

20

Although a variety of cells are available to the art for the expression of proteins, it is most preferred that mammalian cells be utilized. The primary reason for this is predicated upon practicalities, *i.e.*, utilization of, *e.g.*, yeast cells for the expression of a GPCR, while possible,

- 57 -

introduces into the protocol a non-mammalian cell which may not (indeed, in the case of yeast, does not) include the receptor-coupling, genetic-mechanism and secretory pathways that have evolved for mammalian systems – thus, results obtained in non-mammalian cells, while of potential use, are not as preferred as that obtained from mammalian cells. Of the mammalian cells, COS-7, 293 and 293T cells are particularly preferred, although the specific mammalian cell utilized can be predicated upon the particular needs of the artisan.

Unless otherwise noted herein, the following protocol was utilized for the expression of the endogenous and non-endogenous human GPCRs. Table D lists the mammalian cell and number utilized (per 150mm plate) for GPCR expression.

Table D

Receptor Name (Endogenous or Non-Endogenous)	Mammalian Cell (Number Utilized)
GPR17	293 (2×10^4)
GPR30	293 (4×10^4)
APJ	COS-7 (5×10^6)
ETBR-LP2	293 (1×10^7)
	293T (1×10^7)
GHSR	293 (1×10^7)
	293T (1×10^7)
MIG	293 (1×10^7)
Serotonin 5HT _{2A}	293T (1×10^7)
Serotonin 5HT _{2c}	293T (1×10^7)

On day one, mammalian cells were plated out. On day two, two reaction tubes were prepared (the proportions to follow for each tube are per plate): tube A was prepared by mixing 20µg DNA (*e.g.*, pCMV vector; pCMV vector with endogenous receptor cDNA, and pCMV vector with non-endogenous receptor cDNA.) in 1.2ml serum free DMEM (Irvine Scientific,

Irvine, CA); tube B was prepared by mixing 120µl lipofectamine (Gibco BRL) in 1.2ml serum free DMEM. Tubes A and B were then admixed by inversions (several times), followed by incubation at room temperature for 30-45min. The admixture is referred to as the "transfection mixture". Plated cells were washed with 1XPBS, followed by addition of 10ml serum free DMEM. 2.4ml of the transfection mixture was then added to the cells, followed by incubation for 4hrs at 37°C/5% CO₂. The transfection mixture was then removed by aspiration, followed by the addition of 25ml of DMEM/10% Fetal Bovine Serum. Cells were then incubated at 37°C/5% CO₂. After 72hr incubation, cells were then harvested and utilized for analysis.

1. Gi-Coupled Receptors: Co-Transfection with Gs-Coupled Receptors

In the case of GPR30, it has been determined that this receptor couples the G protein Gi. Gi is known to inhibit the enzyme adenylyl cyclase, which is necessary for catalyzing the conversion of ATP to cAMP. Thus, a non-endogenous, constitutively activated form of GPR30 would be expected to be associated with decreased levels of cAMP. Assay confirmation of a non-endogenous, constitutively activated form of GPR30 directly via measurement of decreasing levels of cAMP, while viable, can be preferably measured by cooperative use of a Gs-coupled receptor. For example, a receptor that is Gs-coupled will stimulate adenylyl cyclase, and thus will be associated with an increase in cAMP. The assignee of the present application has discovered that the orphan receptor GPR6 is an endogenous, constitutively activated GPCR. GPR6 couples to the Gs protein. Thus when co-transfected, one can readily verify that a putative GPR30-mutation leads to constitutive activation thereof: *i.e.*, an endogenous, constitutively activated GPR6/endogenous, non-constitutively activated GPR30 cell will evidence an elevated level of cAMP when compared with an endogenous, constitutively active GPR6/non-endogenous, constitutively activated GPR30 (the latter evidencing a comparatively lower level of cAMP).

- 59 -

Assays that detect cAMP can be utilized to determine if a candidate compound is *e.g.*, an inverse agonist to a Gs-associated receptor (*i.e.*, such a compound would decrease the levels of cAMP) or a Gi-associated receptor (or a Go-associated receptor) (*i.e.*, such a candidate compound would increase the levels of cAMP). A variety of approaches known in the art for measuring cAMP can be utilized; a preferred approach relies upon the use of anti-cAMP antibodies. Another approach, and most preferred, utilizes a whole cell second messenger reporter system assay. Promoters on genes drive the expression of the proteins that a particular gene encodes. Cyclic AMP drives gene expression by promoting the binding of a cAMP-responsive DNA binding protein or transcription factor (CREB) which then binds to the promoter at specific sites called cAMP response elements and drives the expression of the gene. Reporter systems can be constructed which have a promoter containing multiple cAMP response elements before the reporter gene, *e.g.*, β -galactosidase or luciferase. Thus, an activated receptor such as GPR6 causes the accumulation of cAMP which then activates the gene and expression of the reporter protein. Most preferably, 293 cells are co-transfected with GPR6 (or another Gs-linked receptor) and GPR30 (or another Gi-linked receptor) plasmids, preferably in a 1:1 ratio, most preferably in a 1:4 ratio. Because GPR6 is an endogenous, constitutively active receptor that stimulates the production of cAMP, GPR6 strongly activates the reporter gene and its expression. The reporter protein such as β -galactosidase or luciferase can then be detected using standard biochemical assays (Chen et al. 1995). Co-transfection of endogenous, constitutively active GPR6 with endogenous, non-constitutively active GPR30 evidences an increase in the luciferase reporter protein. Conversely, co-transfection of endogenous, constitutively active GPR6 with non-endogenous, constitutively active GPR30 evidences a drastic decrease in expression of luciferase. Several reporter plasmids are known and available in the art for measuring a second messenger assay. It is considered well within the

skilled artisan to determine an appropriate reporter plasmid for a particular gene expression based primarily upon the particular need of the artisan. Although a variety of cells are available for expression, mammalian cells are most preferred, and of these types, 293 cells are most preferred. 293 cells were transfected with the reporter plasmid pCRE-Luc/GPR6 and non-endogenous, 5 constitutively activated GPR30 using a Mammalian Transfection™ Kit (Stratagene, #200285) CaPO₄ precipitation protocol according to the manufacturer's instructions (*see*, 28 Genomics 347 (1995) for the published endogenous GPR6 sequence). The precipitate contained 400ng reporter, 80ng CMV-expression plasmid (having a 1:4 GPR6 to endogenous GPR30 or non-endogenous GPR30 ratio) and 20ng CMV-SEAP (a transfection control plasmid encoding secreted alkaline 10 phosphatase). 50% of the precipitate was split into 3 wells of a 96-well tissue culture dish (containing 4X10⁴ cells/well); the remaining 50% was discarded. The following morning, the media was changed. 48 hr after the start of the transfection, cells were lysed and examined for luciferase activity using a Luclite™ Kit (Packard, Cat. # 6016911) and Trilux 1450 Microbeta™ liquid scintillation and luminescence counter (Wallac) as per the vendor's instructions. The data 15 were analyzed using GraphPad Prism 2.0a (GraphPad Software Inc.).

With respect to GPR17, which has also been determined to be Gi-linked, a modification of the foregoing approach was utilized, based upon, *inter alia*, use of another Gs-linked endogenous receptor, GPR3 (*see* 23 Genomics 609 (1994) and 24 Genomics 391 (1994)). Most preferably, 293 cells are utilized. These cells were plated-out on 96 well plates at a density of 2 20 x 10⁴ cells per well and were transfected using Lipofectamine Reagent (BRL) the following day according to manufacturer instructions. A DNA/lipid mixture was prepared for each 6-well transfection as follows: 260ng of plasmid DNA in 100μl of DMEM were gently mixed with 2μl of lipid in 100μl of DMEM (the 260ng of plasmid DNA consisted of 200ng of a 8xCRE-Luc

- 61 -

reporter plasmid (*see below*), 50ng of pCMV comprising endogenous receptor or non-endogenous receptor or pCMV alone, and 10ng of a GPRS expression plasmid (GPRS in pcDNA3 (Invitrogen)). The 8XCRE-Luc reporter plasmid was prepared as follows: vector SRIF- β -gal was obtained by cloning the rat somatostatin promoter (-71/+51) at BglV-HindIII site in the p β gal-
5 Basic Vector (Clontech). Eight (8) copies of cAMP response element were obtained by PCR from an adenovirus template AdpCF126CCRE8 (*see* 7 Human Gene Therapy 1883 (1996)) and cloned into the SRIF- β -gal vector at the Kpn-BglV site, resulting in the 8xCRE- β -gal reporter vector. The 8xCRE-Luc reporter plasmid was generated by replacing the beta-galactosidase gene in the 8xCRE- β -gal reporter vector with the luciferase gene obtained from the pGL3-basic vector
10 (Promega) at the HindIII-BamHI site. Following 30min. incubation at room temperature, the DNA/lipid mixture was diluted with 400 μ l of DMEM and 100 μ l of the diluted mixture was added to each well. 100 μ l of DMEM with 10% FCS were added to each well after a 4hr incubation in a cell culture incubator. The next morning the transfected cells were changed with 200 μ l/well of DMEM with 10% FCS. Eight (8) hours later, the wells were changed to 100 μ l/well of DMEM
15 without phenol red, after one wash with PBS. Luciferase activity were measured the next day using the LucLite™ reporter gene assay kit (Packard) following manufacturer instructions and read on a 1450 MicroBeta™ scintillation and luminescence counter (Wallac).

Figure 4 evidences that constitutively active GPR30 inhibits GPR6-mediated activation of CRE-Luc reporter in 293 cells. Luciferase was measured at about 4.1 relative
20 light units in the expression vector pCMV. Endogenous GPR30 expressed luciferase at about 8.5 relative light units, whereas the non-endogenous, constitutively active GPR30 (L258K), expressed luciferase at about 3.8 and 3.1 relative light units, respectively. Co-transfection of endogenous GPR6 with endogenous GPR30, at a 1:4 ratio, drastically increased luciferase

expression to about 104.1 relative light units. Co-transfection of endogenous GPR6 with non-endogenous GPR30 (L258K), at the same ratio, drastically decreased the expression, which is evident at about 18.2 and 29.5 relative light units, respectively. Similar results were observed with respect to GPR17 with respect to co-transfection with GPR3, as set forth in Figure 5.

Example 3

ASSAYS FOR DETERMINATION OF CONSTITUTIVE ACTIVITY OF NON-ENDOGENOUS GPCRS

A. Membrane Binding Assays

1. [³⁵S]GTPγS Assay

When a G protein-coupled receptor is in its active state, either as a result of ligand binding or constitutive activation, the receptor couples to a G protein and stimulates the release of GDP and subsequent binding of GTP to the G protein. The alpha subunit of the G protein-receptor complex acts as a GTPase and slowly hydrolyzes the GTP to GDP, at which point the receptor normally is deactivated. Constitutively activated receptors continue to exchange GDP for GTP. The non-hydrolyzable GTP analog, [³⁵S]GTPγS, can be utilized to demonstrate enhanced binding of [³⁵S]GTPγS to membranes expressing constitutively activated receptors. The advantage of using [³⁵S]GTPγS binding to measure constitutive activation is that: (a) it is generically applicable to all G protein-coupled receptors; (b) it is proximal at the membrane surface making it less likely to pick-up molecules which affect the intracellular cascade.

The assay utilizes the ability of G protein coupled receptors to stimulate [³⁵S]GTPγS binding to membranes expressing the relevant receptors. The assay can, therefore, be used in the direct identification method to screen candidate compounds to known, orphan and constitutively activated G protein-coupled receptors. The assay is generic and has application

- 63 -

to drug discovery at all G protein-coupled receptors.

The [^{35}S]GTP γ S assay was incubated in 20 mM HEPES and between 1 and about 20mM MgCl_2 (this amount can be adjusted for optimization of results, although 20mM is preferred) pH 7.4, binding buffer with between about 0.3 and about 1.2 nM [^{35}S]GTP γ S (this amount can be adjusted
5 for optimization of results, although 1.2 is preferred) and 12.5 to 75 μg membrane protein (e.g. COS-7 cells expressing the receptor; this amount can be adjusted for optimization, although 75 μg is preferred) and 1 μM GDP (this amount can be changed for optimization) for 1 hour. Wheatgerm agglutinin beads (25 μl ; Amersham) were then added and the mixture was incubated for another 30 minutes at room temperature. The tubes were then centrifuged at 1500 x g for 5
10 minutes at room temperature and then counted in a scintillation counter.

A less costly but equally applicable alternative has been identified which also meets the needs of large scale screening. Flash platesTM and WallacTM scintistrips may be utilized to format a high throughput [^{35}S]GTP γ S binding assay. Furthermore, using this technique, the assay can be utilized for known GPCRs to simultaneously monitor tritiated ligand binding to the receptor at the
15 same time as monitoring the efficacy via [^{35}S]GTP γ S binding. This is possible because the Wallac beta counter can switch energy windows to look at both tritium and ^{35}S -labeled probes. This assay may also be used to detect other types of membrane activation events resulting in receptor activation. For example, the assay may be used to monitor ^{32}P phosphorylation of a variety of receptors (both G protein coupled and tyrosine kinase receptors). When the membranes are
20 centrifuged to the bottom of the well, the bound [^{35}S]GTP γ S or the ^{32}P -phosphorylated receptor will activate the scintillant which is coated of the wells. Scinti[®] strips (Wallac) have been used to demonstrate this principle. In addition, the assay also has utility for measuring ligand binding to receptors using radioactively labeled ligands. In a similar manner, when the radiolabeled bound

ligand is centrifuged to the bottom of the well, the scintistrip label comes into proximity with the radiolabeled ligand resulting in activation and detection.

Representative results of graph comparing Control (pCMV), Endogenous APJ and Non-Endogenous APJ, based upon the foregoing protocol, are set forth in Figure 6.

5 2. Adenylyl Cyclase

A Flash Plate™ Adenylyl Cyclase kit (New England Nuclear; Cat. No. SMP004A) designed for cell-based assays was modified for use with crude plasma membranes. The Flash Plate wells contain a scintillant coating which also contains a specific antibody recognizing cAMP. The cAMP generated in the wells was quantitated by a direct competition for binding of
10 radioactive cAMP tracer to the cAMP antibody. The following serves as a brief protocol for the measurement of changes in cAMP levels in membranes that express the receptors.

Transfected cells were harvested approximately three days after transfection. Membranes were prepared by homogenization of suspended cells in buffer containing 20mM HEPES, pH 7.4 and 10mM MgCl₂. Homogenization was performed on ice using a Brinkman Polytron™ for
15 approximately 10 seconds. The resulting homogenate was centrifuged at 49,000 X g for 15 minutes at 4°C. The resulting pellet was then resuspended in buffer containing 20mM HEPES, pH 7.4 and 0.1 mM EDTA, homogenized for 10 seconds, followed by centrifugation at 49,000 X g for 15 minutes at 4°C. The resulting pellet can be stored at -80°C until utilized. On the day of measurement, the membrane pellet was slowly thawed at room temperature, resuspended in buffer.
20 containing 20mM HEPES, pH 7.4 and 10mM MgCl₂ (these amounts can be optimized, although the values listed herein are preferred), to yield a final protein concentration of 0.60mg/ml (the resuspended membranes were placed on ice until use).

cAMP standards and Detection Buffer (comprising 2 µCi of tracer [¹²⁵I cAMP (100 µl) to

- 65 -

11 ml Detection Buffer) were prepared and maintained in accordance with the manufacturer's instructions. Assay Buffer was prepared fresh for screening and contained 20mM HEPES, pH 7.4, 10mM MgCl₂, 20mM (Sigma), 0.1 units/ml creatine phosphokinase (Sigma), 50 μM GTP (Sigma), and 0.2 mM ATP (Sigma); Assay Buffer can be stored on ice until utilized. The assay
5 was initiated by addition of 50ul of assay buffer followed by addition of 50ul of membrane suspension to the NEN Flash Plate. The resultant assay mixture is incubated for 60 minutes at room temperature followed by addition of 100ul of detection buffer. Plates are then incubated an additional 2-4 hours followed by counting in a Wallac MicroBeta scintillation counter. Values of cAMP/well are extrapolated from a standard cAMP curve which is contained within each assay
10 plate. The foregoing assay was utilized with respect to analysis of MIG.

B. Reporter-Based Assays

1. CREB Reporter Assay (Gs-associated receptors)

A method to detect Gs stimulation depends on the known property of the transcription factor CREB, which is activated in a cAMP-dependent manner. A PathDetect CREB trans-
15 Reporting System (Stratagene, Catalogue # 219010) was utilized to assay for Gs coupled activity in 293 or 293T cells. Cells were transfected with the plasmids components of this above system and the indicated expression plasmid encoding endogenous or mutant receptor using a Mammalian Transfection Kit (Stratagene, Catalogue #200285) according to the manufacturer's instructions. Briefly, 400 ng pFR-Luc (luciferase reporter plasmid containing
20 Gal4 recognition sequences), 40 ng pFA2-CREB (Gal4-CREB fusion protein containing the Gal4 DNA-binding domain), 80 ng CMV-receptor expression plasmid (comprising the receptor) and 20 ng CMV-SEAP (secreted alkaline phosphatase expression plasmid; alkaline phosphatase activity is measured in the media of transfected cells to control for variations in

transfection efficiency between samples) were combined in a calcium phosphate precipitate as per the Kit's instructions. Half of the precipitate was equally distributed over 3 wells in a 96-well plate, kept on the cells overnight, and replaced with fresh medium the following morning. Forty-eight (48) hr after the start of the transfection, cells were treated and assayed for luciferase activity as set forth with respect to the GPR30 system, above. This assay was used with respect to GHSR.

2. AP1 reporter assay (Gq-associated receptors)

A method to detect Gq stimulation depends on the known property of Gq-dependent phospholipase C to cause the activation of genes containing AP1 elements in their promoter. A Pathdetect AP-1 cis-Reporting System (Stratagene, Catalogue # 219073) was utilized following the protocol set forth above with respect to the CREB reporter assay, except that the components of the calcium phosphate precipitate were 410 ng pAP1-Luc, 80 ng receptor expression plasmid, and 20 ng CMV-SEAP. This assay was used with respect to ETBR-LP2

C. Intracellular IP3 Accumulation Assay

On day 1, cells comprising the serotonin receptors (endogenous and mutated) were plated onto 24 well plates, usually 1×10^5 cells/well. On day 2 cells were transfected by firstly mixing 0.25ug DNA in 50 ul serumfree DMEM/well and 2 ul lipofectamine in 50 μ l serumfree DMEM/well. The solutions were gently mixed and incubated for 15-30 min at room temperature. Cells were washed with 0.5 ml PBS and 400 μ l of serum free media was mixed with the transfection media and added to the cells. The cells were then incubated for 3-4 hrs at $37^\circ\text{C}/5\%\text{CO}_2$ and then the transfection media was removed and replaced with 1ml/well of regular growth media. On day 3 the cells were labeled with ^3H -myo-inositol. Briefly, the media was removed the cells were washed with 0.5 ml PBS. Then 0.5 ml inositol-free/serumfree media (GIBCO BRL) was added/well with 0.25 μCi of ^3H -myo-inositol / well

- 67 -

and the cells were incubated for 16-18 hrs o/n at 37°C/5%CO₂. On Day 4 the cells were washed with 0.5 ml PBS and 0.45 ml of assay medium was added containing inositol-free/serum free media 10 µM pargyline 10 mM lithium chloride or 0.4 ml of assay medium and 50 ul of 10x ketanserin (ket) to final concentration of 10µM. The cells were then
5 incubated for 30 min at 37°C. The cells were then washed with 0.5 ml PBS and 200 ul of fresh/icecold stop solution (1M KOH; 18 mM Na-borate; 3.8 mM EDTA) was added/well. The solution was kept on ice for 5-10 min or until cells were lysed and then neutralized by 200 µl of fresh/ice cold neutralization sol. (7.5 % HCL). The lysate was then transferred into 1.5 ml eppendorf tubes and 1 ml of chloroform/methanol (1:2) was added/tube. The solution
10 was vortexed for 15 sec and the upper phase was applied to a Biorad AG1-X8 anion exchange resin (100-200 mesh). Firstly, the resin was washed with water at 1:1.25 W/V and 0.9 ml of upper phase was loaded onto the column. The column was washed with 10 mls of 5 mM myo-inositol and 10 ml of 5 mM Na-borate/60mM Na-formate. The inositol tris phosphates were eluted into scintillation vials containing 10 ml of scintillation cocktail with
15 2 ml of 0.1 M formic acid/ 1 M ammonium formate. The columns were regenerated by washing with 10 ml of 0.1 M formic acid/3M ammonium formate and rinsed twice with dd H₂O and stored at 4°C in water.

Figure 7 provides an illustration of IP₃ production from the human 5-HT_{2A} receptor that incorporates the C322K mutation. While these results evidence that the Proline Mutation
20 Algorithm approach constitutively activates this receptor, for purposes of using such a receptor for screening for identification of potential therapeutics, a more robust difference would be preferred. However, because the activated receptor can be utilized for understanding and elucidating the role of constitutive activation and for the identification of compounds that

can be further examined, we believe that this difference is itself useful in differentiating between the endogenous and non-endogenous versions of the human 5HT_{2A} receptor.

D. Result Summary

The results for the GPCRs tested are set forth in Table E where the Per-Cent Increase indicates the percentage difference in results observed for the non-endogenous GPCR as compared to the endogenous GPCR; these values are followed by parenthetical indications as to the type of assay utilized. Additionally, the assay system utilized is parenthetically listed (and, in cases where different Host Cells were used, both are listed). As these results indicate, a variety of assays can be utilized to determine constitutive activity of the non-endogenous versions of the human GPCRs.

Those skilled in the art, based upon the foregoing and with reference to information available to the art, are credited with the ability to select and/or maximize a particular assay approach that suits the particular needs of the investigator.

Table E

Receptor Identifier (Codon Mutation)	Per-Cent Difference
GPR17 (V234K)	74.5 (CRE-Luc)
GPR30 (L258K)	71.6 (CREB)
APJ (L247K)	49.0 (GTP γ S)
ETBR-LP2 (N358K)	48.4(AP1-Luc - 293) 61.1(AP1-Luc - 293T)
GHSR (V262K)	58.9(CREB - 293) 35.6(CREB - 293T)

- 69 -

MIG (I230K)	39 (cAMP)
Serotonin 5HT _{2A} (C322K)	33.2 (IP ₃)
Serotonin 5HT _{2C} (S310K)	39.1 (IP ₃)

Example 6**Tissue Distribution of Endogenous Orphan GPCRs**

Using a commercially available human-tissue dot-blot format, endogenous orphan GPCRs were probed for a determination of the areas where such receptors are localized. Except as indicate below, the entire receptor cDNA (radiolabelled) was used as the probe: radiolabeled probe was generated using the complete receptor cDNA (excised from the vector) using a Prime-It II™ Random Primer Labeling Kit (Stratagene, #300385), according to manufacturer's instructions. A human RNA Master Blot™ (Clontech, #7770-1) was hybridized with the GPCR radiolabeled probe and washed under stringent conditions according manufacturer's instructions. The blot was exposed to Kodak BioMax Autoradiography film overnight at -80°C.

Representative dot-blot format results are presented in Figure 8 for GPR1 (8A), GPR30 (8B), and APJ (8C), with results being summarized for all receptors in Table F

Table F

GPCR	Tissue Distribution (highest levels, relative to other tissues in the dot-blot)
GPR1	Placenta, Ovary, Adrenal

	GPR4	Broad; highest in Heart, Lung, Adrenal, Thyroid, Spinal Cord
	GPR5	Placenta, Thymus, Fetal Thymus Lesser levels in spleen, fetal spleen
	GPR7	Liver, Spleen, Spinal Cord, Placenta
	GPR8	No expression detected
5	GPR9-6	Thymus, Fetal Thymus Lesser levels in Small Intestine
	GPR18	Spleen, Lymph Node, Fetal Spleen, Testis
	GPR20	Broad
	GPR21	Broad; very low abundance
	GPR22	Heart, Fetal Heart Lesser levels in Brain
10	GPR30	Stomach
	GPR31	Broad
	BLR1	Spleen
	CEPR	Stomach, Liver, Thyroid, Putamen
	EBI1	Pancreas Lesser levels in Lymphoid Tissues
15	EBI2	Lymphoid Tissues, Aorta, Lung, Spinal Cord
	ETBR-LP2	Broad; Brain Tissue
	GPCR-CNS	Brain Lesser levels in Testis, Placenta
	GPR-NGA	Pituitary Lesser levels in Brain
20	H9	Pituitary
	HB954	Aorta, Cerebellum Lesser levels in most other tissues
	HM74	Spleen, Leukocytes, Bone marrow, Mammary Glands, Lung, Trachea
	MIG	Low levels in Kidney, Liver, Pancreas, Lung, Spleen
	ORG1	Pituitary, Stomach, Placenta
	V28	Brain, Spleen, Peripheral Leukocytes

25 Based upon the foregoing information, it is noted that human GPCRs can also be assessed for distribution in diseased tissue; comparative assessments between "normal" and diseased tissue can then be utilized to determine the potential for over-expression or under-expression of a particular receptor in a diseased state. In those circumstances where it is desirable to utilize the non-endogenous versions of the human GPCRs for the purpose of screening to directly identify

- 71 -

candidate compounds of potential therapeutic relevance, it is noted that inverse agonists are useful in the treatment of diseases and disorders where a particular human GPCR is over-expressed, whereas agonists or partial agonists are useful in the treatment of diseases and disorders where a particular human GPCR is under-expressed.

5 As desired, more detailed, cellular localization of the receptors, using techniques well-known to those in the art (*e.g.*, in-situ hybridization) can be utilized to identify particular cells within these tissues where the receptor of interest is expressed.

It is intended that each of the patents, applications, and printed publications mentioned in this patent document be hereby incorporated by reference in their entirety.

10 As those skilled in the art will appreciate, numerous changes and modifications may be made to the preferred embodiments of the invention without departing from the spirit of the invention. It is intended that all such variations fall within the scope of the invention.

Although a variety of expression vectors are available to those in the art, for purposes of utilization for both the endogenous and non-endogenous human GPCRs, it is most preferred that
15 the vector utilized be pCMV. This vector has been deposited with the American Type Culture Collection (ATCC) on October 13, 1998 (10801 University Blvd., Manassas, VA 20110-2209 USA) under the provisions of the Budapest Treaty for the International Recognition of the Deposit of Microorganisms for the Purpose of patent Procedure. The vector was tested by the ATCC on _____, 1998 and determined to be viable on _____, 1998. The ATCC has assigned
20 the following deposit number to pCMV: _____.

CLAIMS

What is claimed is:

1. A constitutively active, non-endogenous version of an endogenous human orphan G protein-coupled receptor (GPCR) comprising the following amino acid residues (carboxy-terminus to amino-terminus orientation) transversing the transmembrane-6 (TM6) and intracellular loop-3 (IC3) regions of the non-endogenous GPCR:

$P^I AA_{15} X$

wherein:

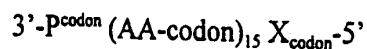
- (1) P^I is an amino acid residue located within the TM6 region of the non-endogenous GPCR, where P^I is selected from the group consisting of (i) the endogenous orphan GPCR proline residue, and (ii) a non-endogenous amino acid residue other than proline;
- (2) AA_{15} are 15 amino acid residues selected from the group consisting of (a) the 15 endogenous amino acid residues of the endogenous orphan GPCR, (b) 15 non-endogenous amino acid residues, and (c) a combination of 15 amino acid residues, the combination comprising at least one endogenous amino acid residue of the endogenous orphan GPCR and at least one non-endogenous amino acid residue, excepting that none of the 15 endogenous amino acid residues that are positioned within the TM6 region of the GPCR is proline; and
- (2) X is a non-endogenous amino acid residue located within the IC3 region of said non-endogenous GPCR.
2. The non-endogenous human GPCR of claim 1 wherein P^I is the endogenous proline

- 73 -

residue.

3. The non-endogenous human GPCR of claim 1 wherein P¹ is a non-endogenous amino acid residue other than a proline residue.
4. The non-endogenous human GPCR of claim 1 wherein AA₁₅ are the 15 endogenous amino acid residues of the endogenous GPCR.
5. The non-endogenous human GPCR of claim 1 wherein X is selected from the group consisting of lysine, histidine, arginine and alanine residues, excepting that when the endogenous amino acid in position X of said endogenous human GPCR is lysine, X is selected from the group consisting of histidine, arginine and alanine.
- 10 6. The non-endogenous human GPCR of claim 1 wherein X is a lysine residue, excepting that when the endogenous amino acid in position X of said endogenous human GPCR is lysine, X is an amino acid other than lysine.
7. The non-endogenous human GPCR of claim 4 wherein X is a lysine residue, excepting that when the endogenous amino acid in position X of said endogenous human GPCR is lysine, X is an amino acid other than lysine.
- 15 8. The non-endogenous, human GPCR of claim 1 wherein P¹ is a proline residue and X is a lysine residue, excepting that when the endogenous amino acid in position X of said endogenous human GPCR is lysine, X is an amino acid other than lysine.
9. A host cell comprising the non-endogenous human GPCR of claim 1.
- 20 10. The material of claim 9 wherein said host cell is of mammalian origin.
11. The non-endogenous human GPCR of claim 1 in a purified and isolated form.
12. A nucleic acid sequence encoding a constitutively active, non-endogenous version of an endogenous human orphan G protein-coupled receptor (GPCR) comprising the following

nucleic acid sequence region transversing the transmembrane-6 (TM6) and intracellular loop-3 (IC3) regions of the orphan GPCR:



wherein:

- 5 (1) P_{codon} is a nucleic acid encoding region within the TM6 region of the non-endogenous GPCR, where P_{codon} encodes an amino acid selected from the group consisting of (i) the endogenous GPCR proline residue, and (ii) a non-endogenous amino acid residue other than proline;
 - 10 (2) $(\text{AA-codon})_{15}$ are 15 codons encoding 15 amino acid residues selected from the group consisting of (a) the 15 endogenous amino acid residues of the endogenous orphan GPCR; (b) 15 non-endogenous amino acid residues, and (c) a combination of 15 amino acid residues, the combination comprising at least one endogenous amino acid residue of the endogenous orphan GPCR and at least one non-
15 endogenous amino acid residue, excepting that none of the 15 endogenous amino acid residues that are positioned within the TM6 region of the orphan GPCR is proline; and
 - 20 (3) X_{codon} is a nucleic acid encoding region residue located within the IC3 region of said non-endogenous human GPCR, where X_{codon} encodes a non-endogenous amino acid.
13. The nucleic acid sequence of claim 12 wherein P_{codon} encodes an endogenous proline residue.
 14. The nucleic acid sequence of claim 12 wherein P_{codon} encodes a non-endogenous

- 75 -

- amino acid residue other than a proline residue.
15. The nucleic acid sequence of claim 12 wherein X_{codon} encodes a non-endogenous amino acid selected from the group consisting of lysine, histidine, arginine and alanine, excepting that when the endogenous amino acid in position X of said
- 5 endogenous human GPCR is lysine, X_{codon} encodes an amino acid selected from the group consisting of histidine, arginine and alanine.
16. The nucleic acid sequence of claim 13 wherein X_{codon} encodes a non-endogenous lysine amino acid excepting that when the endogenous amino acid in position X of said endogenous human GPCR is lysine, X_{codon} encodes an amino acid selected from
- 10 the group consisting of histidine, arginine and alanine.
17. The nucleic acid sequence of claim 12 wherein X_{codon} is selected from the group consisting of AAA, AAG, GCA, GCG, GCC and GCU.
18. The nucleic acid sequence of claim 12 wherein X_{codon} is selected from the group consisting of AAA and AAG.
- 15 19. The nucleic acid sequence of claim 12 wherein P_{codon} is selected from the group consisting of CCA, CCC, CCG and CCU, and X_{codon} is selected from the group consisting of AAA and AAG.
20. A vector comprising the nucleic acid sequence of claim 12.
21. A plasmid comprising the nucleic acid sequence of claim 12.
- 20 22. A host cell comprising the nucleic acid sequence of claim 21.
23. The nucleic acid sequence of claim 12 in a purified and isolated form.
24. A method for selecting for alteration an endogenous amino acid residue within the third intracellular loop of a human G protein-coupled receptor ("GPCR"), said receptor

comprising a transmembrane 6 region and an intracellular loop 3 region, which endogenous amino acid, when altered to a non-endogenous amino acid, constitutively activates said human GPCR, comprising the following steps:

- 5 (a) identifying an endogenous proline residue within the transmembrane 6 region of a human GPCR;
- (b) identifying, by moving in a direction of the carboxy-terminus region of said GPCR towards the amino-terminus region of said GPCR, the endogenous, 16th amino acid residue from said proline residue;
- 10 (c) altering the endogenous residue of step (b) to a non-endogenous amino acid residue to create a non-endogenous version of an endogenous human GPCR; and
- (d) determining whether the non-endogenous human GPCR of step (c) is constitutively active.

25. The method of claim 24 wherein the amino acid residue that is two residues from said
15 proline residue in the transmembrane 6 region, in a carboxy-terminus to amino-terminus direction, is tryptophan.

26. A constitutively active, non-endogenous human GPCR produced by the process of claim 24.

27. A constitutively active, non-endogenous human GPCR produced by the process of
20 claim 25.

28. An algorithmic approach for creating a non-endogenous, constitutively active version of an endogenous human G protein coupled receptor (GPCR), said endogenous GPCR comprising a transmembrane 6 region and an intracellular loop 3 region, the

- 77 -

algorithmic approach comprising the steps of:

- (a) selecting an endogenous human GPCR comprising a proline residue in the transmembrane-6 region;
 - (b) identifying, by counting 16 amino acid residues from the proline residue of step (a), in a carboxy-terminus to amino-terminus direction, an endogenous amino acid residue;
 - (c) altering the identified amino acid residue of step (b) to a non-endogenous amino acid residue to create a non-endogenous version of the endogenous human GPCR; and
 - (d) determining if the non-endogenous version of the endogenous human GPCR of step (c) is constitutively active.
29. The algorithmic approach of claim 28 wherein the amino acid residue that is two residues from said proline residue in the transmembrane 6 region, in a carboxy-terminus to amino-terminus direction, is tryptophan.
30. A constitutively active, non-endogenous human GPCR produced by the algorithmic approach of claim 28.
31. A constitutively active, non-endogenous human GPCR produced by the algorithmic approach of claim 29.
32. A method for directly identifying a compound selected from the group consisting of inverse agonists, agonists and partial agonists to a non-endogenous, constitutively activated human G protein coupled receptor, said receptor comprising a transmembrane-6 region and an intracellular loop-3 region, comprising the steps of:
- (a) selecting an endogenous human GPCR;

- (b) identifying a proline residue within the transmembrane-6 region of the GPCR of step (a);
 - (c) identifying, in a carboxy-terminus to amino-terminus direction, the endogenous, 16th amino acid residue from the proline residue of step (b);
 - 5 (d) altering the endogenous amino acid of step (c) to a non-endogenous amino acid;
 - (e) confirming that the non-endogenous GPCR of step (d) is constitutively active;
 - (f) contacting a candidate compound with the non-endogenous, constitutively-activated GPCR of step (e); and
 - 10 (g) determining, by measurement of the compound efficacy at said contacted receptor, whether said compound is an inverse agonist, agonist or partial agonist of said receptor.
-
- 33. The method of claim 32 wherein the non-endogenous amino acid of step (d) is lysine.
 - 34. A compound directly identified by the method of claim 32.
 - 15 35. The method of claim 32 wherein the directly identified compound is an inverse agonist.
 - 36. The method of claim 32 wherein the directly identified compound is an agonist.--
 - 37. The method of claim 32 wherein the directly identified compound is a partial agonist.
 - 38. A composition comprising the inverse agonist of claim 35.
 - 20 39. A composition comprising the agonist of claim 36.
 - 40. A composition comprising the partial agonist of claim 37.
 - 41. A method for directly identifying an inverse agonist to a non-endogenous,

- 79 -

constitutively activated human G protein coupled receptor ("GPCR"), said GPCR comprising a transmembrane-6 region and an intracellular loop-3 region, comprising the steps of:

- (a) selecting an endogenous human GPCR;
 - (b) identifying a proline residue within the transmembrane-6 region of the GPCR of
5 step (a);
 - (c) identifying, in a carboxy-terminus to amino-terminus direction, the
 endogenous, 16th amino acid residue from the proline residue of step (b);
 - (d) altering the endogenous amino acid of step (c) to a non-endogenous lysine residue;
 - (e) confirming that the non-endogenous GPCR of step (d) is constitutively active;
 - 10 (f) contacting a candidate compound with the non-endogenous, constitutively-
 activated GPCR of step (e); and
 - (g) determining, by measurement of the compound efficacy at said contacted receptor,
 whether said compound is an inverse agonist of said receptor.
42. An inverse agonist directly identified by the method of claim 37.
- 15 43. A composition comprising an inverse agonist of claim 38.

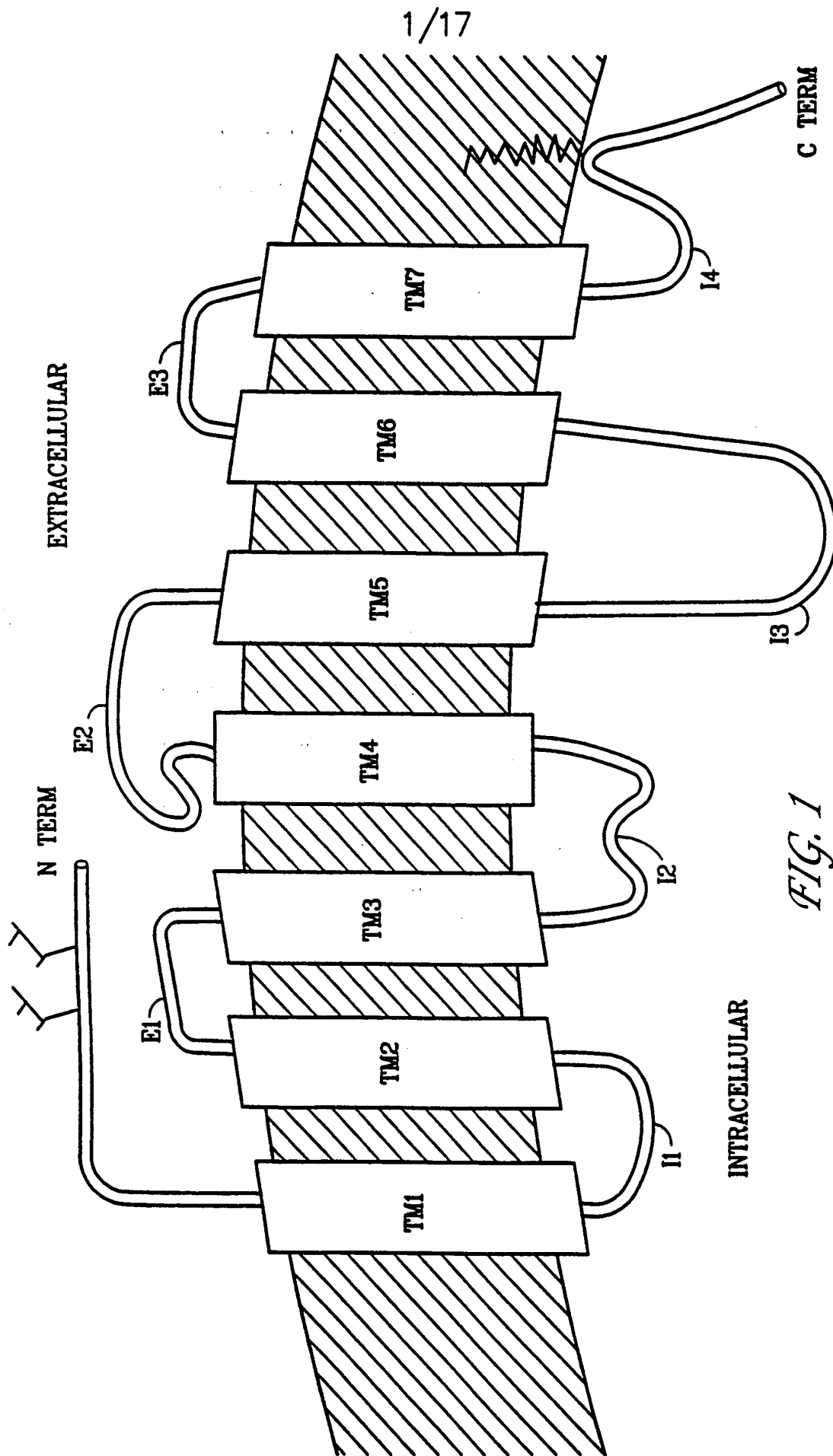


FIG. 1

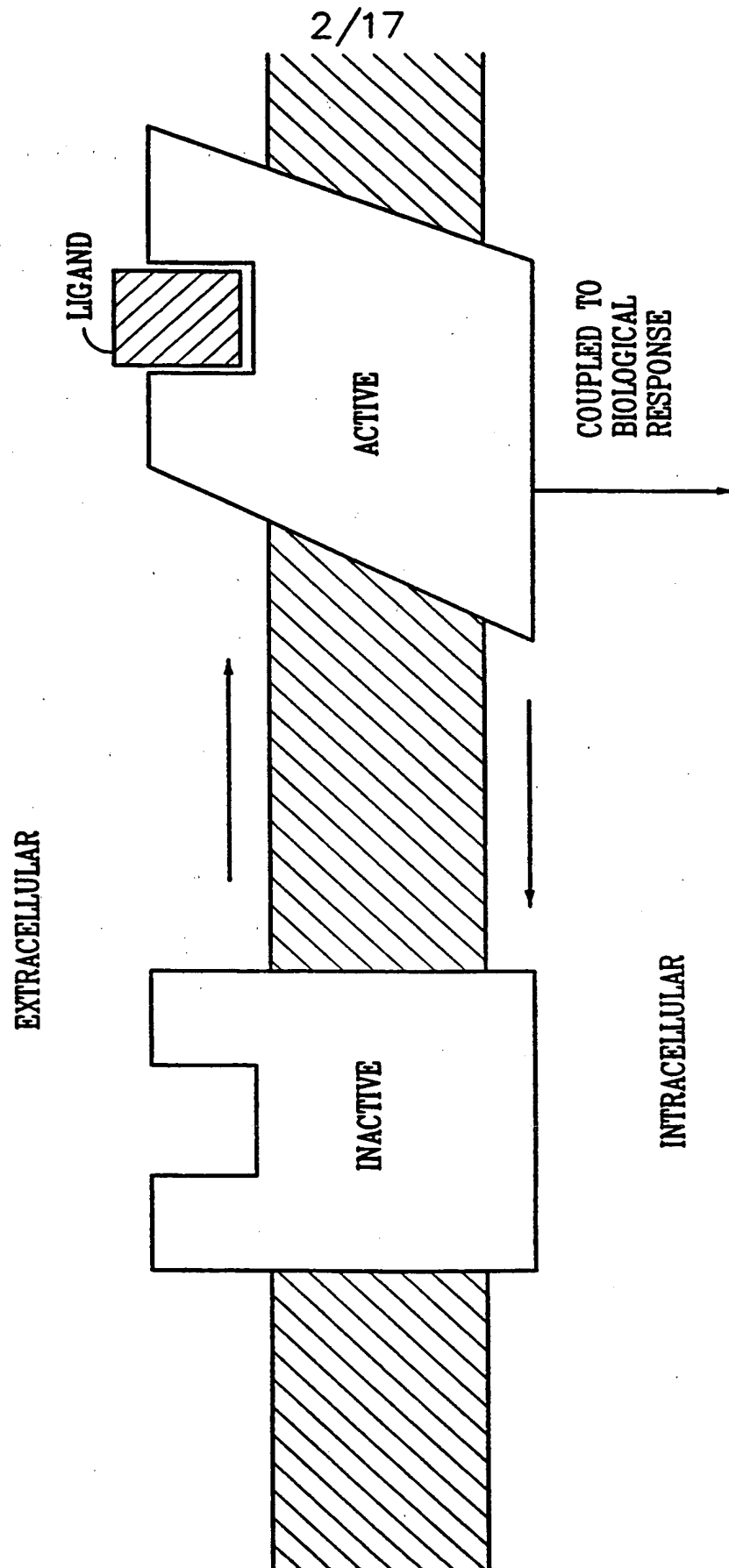


FIG. 2

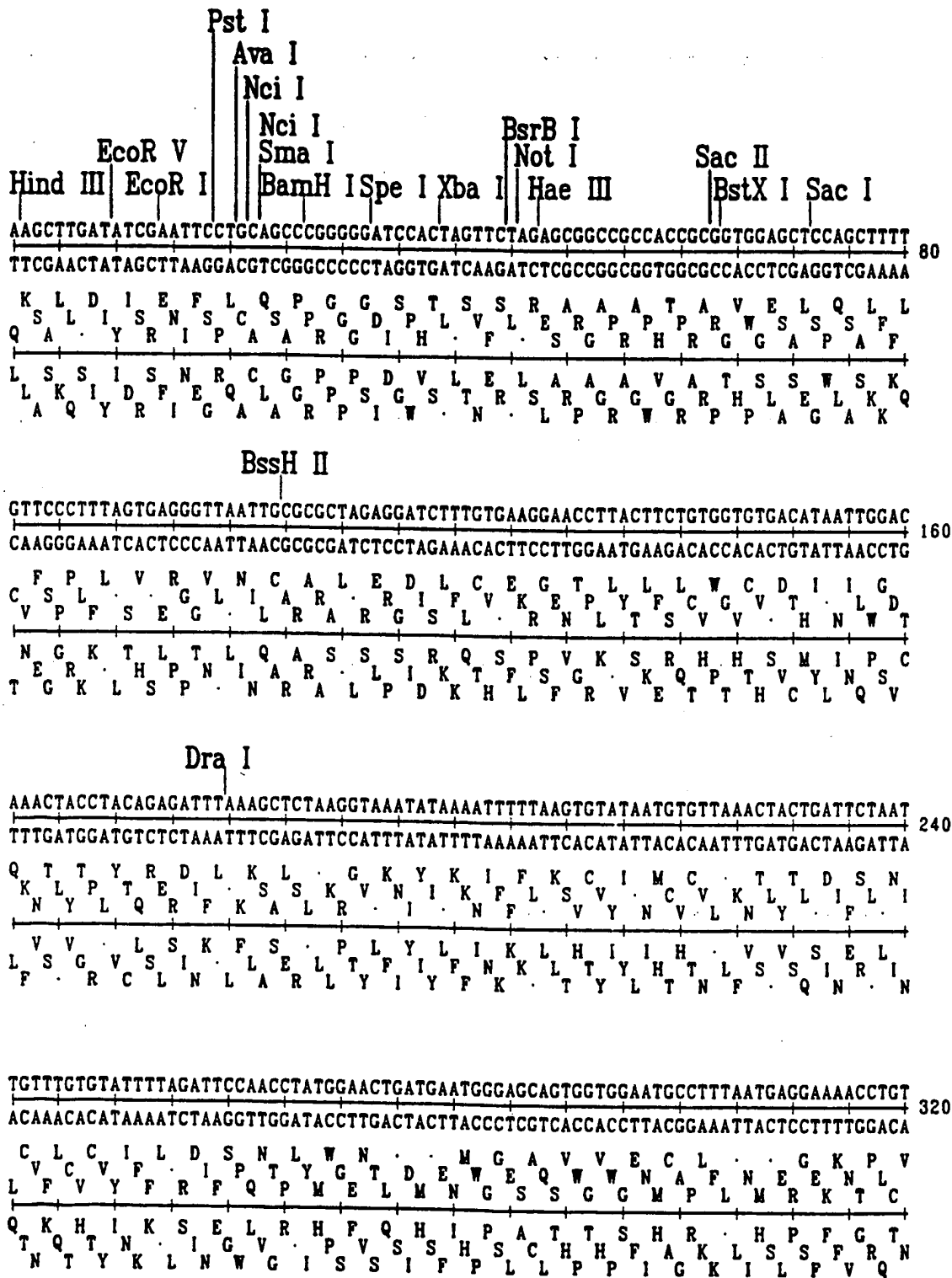


FIG. 3A

4/17

TTTGCTCAGAAGAAATGCCATCTAGTGATGATGAGGCTACTGCTGACTCTCAACATTCTACTCCTCCAAAAAGAAGAGA 400
 AAACGAGTCTTCTTTACGGTAGATCACTACTACTCCGATGACGACTGAGAGTTGTAAGATGAGGAGGTTTTTCTTCTCT
 L L R R N A I . . . G Y C . L S T F Y S S K K E E
 F C A S Q K K C H L V M M R A L L T L N I L L L Q K R R E
 K S L L F A M . H H H P . Q Q S E V N . E E L F S S L F
 K A . F F H W R T I I L S S S V R L M R S R W F L L S

Sty I

AAGGTAGAAGACCCCAAGGACTTTTCCTTCAGAATTGCTAAGTTTTTTGAGTCATGCTGTGTTTAGTAATAGAACTCTTGC 480
 TTCCATCTTCTGGGGTTCCTGAAAGGAAGTCTTAACGATTCAAAAACTCAGTACGACACAAATCATTATCTTGAGAACC
 K G R R P Q G L S F R I A K F F E S C V . N S C A
 K V E D P K D F P S E L L S F L S V M L C L V I E L L A
 P L L G W P S E K L I A L N K S D H Q T . Y Y F E Q
 F T S S G L S K G E S N S L K K L . A T N L L L V R A
 L Y F V G L V K R . F Q . T K Q T M S H K T I S S K S

TTGCTTTGCTATTTACACCACAAAGGAAAAAGCTGCACTGCTATACAAGAAAATTATGGAAAAATATTCTGTAACCTTTA 560
 AACGAAACGATAAATGTGGTGTTTCCTTTTCGACGTGACGATATGTTCTTTTAATACCTTTTTATAAGACATTGGAAT
 L L C Y L H H K G K S C T A I Q E N Y G K I F C N L Y
 L C F A I Y T T Q K E K A A L L Y T K K I M E K Y S V T P L
 L A L L F T P Q R K K L H C Y T R K L W K N I L . P L
 K S Q . K C V W L P F F L Q V A I C S F . P F I N Q L R K I
 A K S N V G C L F F S C Q . V L F N H F F I R Y G K

AseI

TAAGTAGGCATAACAGTTATAATCATAACTGTTTTTCTTACTCCACACAGGCATAGAGTGTCTGCTATTAATAAC 640
 ATTCATCCGTATTGTCAATATTAGTATTGTATGACAAAAAGAATGAGGTGTGTCGGTATCTCACAGACGATAATTATTG
 I K . A . Q L . S . H T V F S Y S T Q A . S V C Y . N
 I S R H N S Y N H N I L F F L T P H R H R V S A I N N
 I V G I T V I I I T Y C F F L L H T G I E C L L L I T
 L Y A Y C N Y D Y C V T K E . E V C A Y L T Q . Y S
 L L C L L L M V S N K K R V E C L C L T D A I L L V
 Y T P M V T I I M V Y Q K K K S W V P M S H R S N I V

Rsa I

TATGCTCAAAAATTGTGTACCTTTAGCTTTTTAATTTGTAAAGGGGTTAATAAGGAATATTTGATGTATAGTCCTTGAC 720
 ATACGAGTTTTTAACACATGGAATCGAAAAATTAACATTTCCCAATTATTCCTTATAAACTACATATCACGGAAC TG
 L C S K I V Y L . L F N L . R G . G I F D Y . C L D
 Y A Q K L C T P L S F L I C K G V N K E Y L M Y S A L T
 M L K N C V P L A F . F V K G L I R N I . C I V P .
 H E F I T Y R . S K L K Y L P . Y P I N S T Y H R S V
 I A S F N H V K L K K I Q L P T L L S Y K I Y L A K Q S

FIG. 3B

6/17

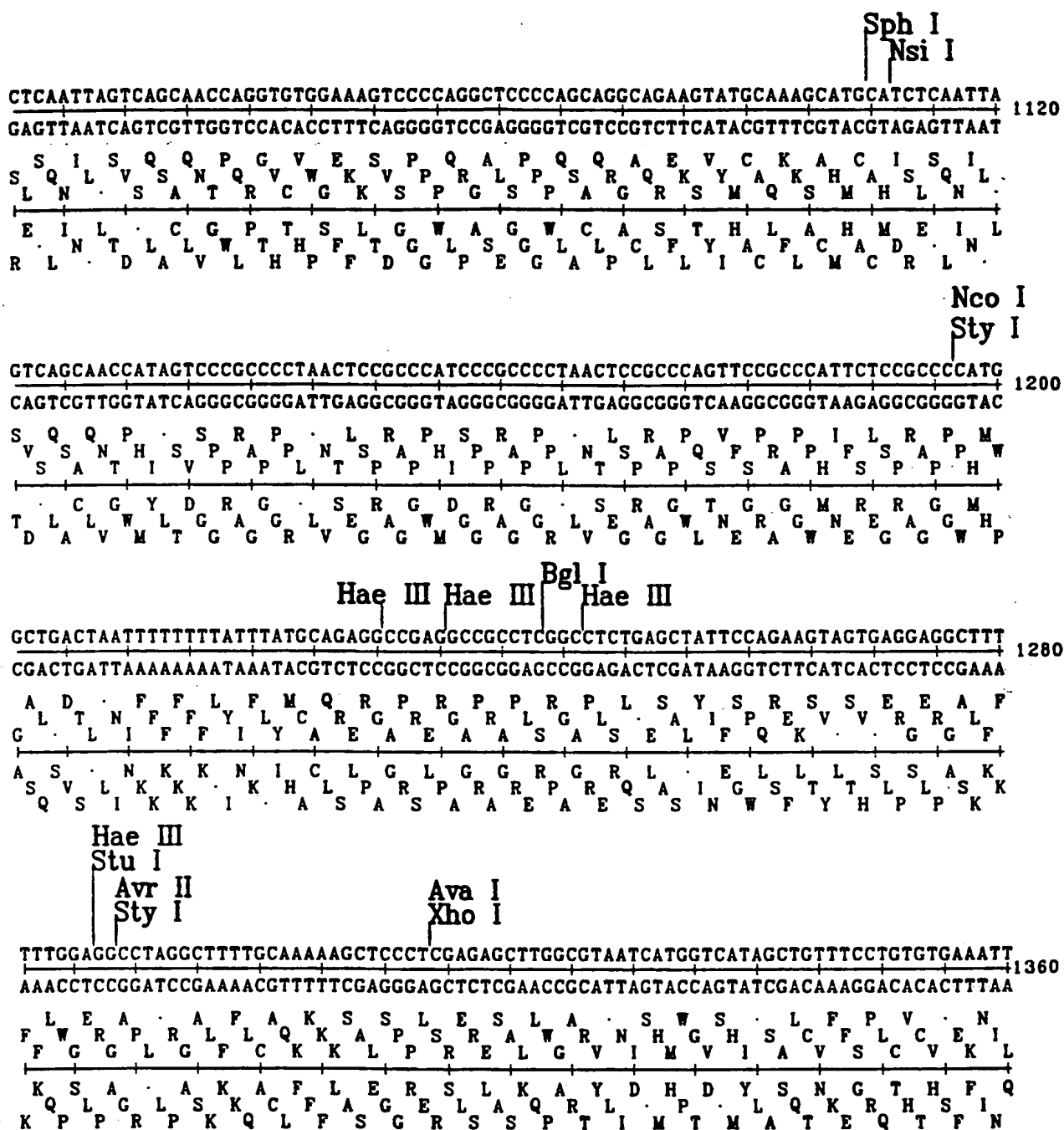


FIG. 3D

Bsrb I

GTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCTAATGAGTGAGCTAA 1440
 CAATAGGCGAGTGTTAAGGTGTGTTGTATGCTCGGCCTTCGTATTTACATTTCGGACCCACGGATTACTCACTCGATT
 C Y P L T I P H N I R A G S I K C K A W G A V S
 V I R S Q F H T T Y E P E A S V K P G V P N E A N
 L S A H N S T Q H T S R K H K V S L G C L M S E L
 G S V I G C L M R A P L M F H L A Q P A H T L
 T I R E C N W V V Y S G S A Y L T F G P T G L S H A I
 N D A L E V C C V L R F C L T Y L R P H R I L S S V

AseI

Pvu II AseI Hae III

CTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGTCGATTAATGAATCGGCCA 1520
 GAGTGTAATTAACGCAACGCGAGTGACGGGCGAAAGGTCAGCCCTTTGGACAGCAGGTCGACGTAATTACTTAGCCGGT
 L T L I A L R S L P A F Q S G N L S C Q L H I G Q
 S H L L R C A H C P L S S R E T C R A S C I N E S A
 T H I N C V A L T A R F P V G K P V V P A A L M N R P
 S V N I A N R E S G A K W D P F R D H W S C H I P W
 E C M L N R Q A Q G S E L R S V Q R A L Q M L S D A L
 M L Q T A S V A R K G T P F G T T G A A N I F R G

Sap I

ACGCGCGGGGAGAGGCGGTTTGGCTATTGGGCGCTCTTCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTGCTTCGGC 1600
 TGCGCGCCCTCTCGGCCAAACGCATAACCCCGGAGAAGGCGAAGGAGCGAGTGACTGAGCGACGCGAGCCAGCAAGCCC
 R A G R G G L R I G R S S A S S L T D S L R S V V R
 N A R G E A V C V L G A L P L P R S L T R C A R S F G
 T R G E R R F A Y W A L F R F L A H L A A L G R S A
 R A P L P P K R I P R E E A E E S V S E S R E T T R S
 A R P S A T Q T N P A R G S G R E S V R Q A R D N P
 V R P S L R N A Y Q A S K R K R A Q S A A S P R E A

BsrB I

TGCGCGGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATG 1680
 ACGCGGCTCGCCATAGTCGAGTGAGTTTCGCCATTATGCCAATAGGTGTCTTAGTCCCCTATTGCGTCCTTTCTTGTA
 L R R A V S A H S K A V I R L S T E S G D N A G K N M
 C G E R Y Q L T Q R R Y G Y P Q N Q G I T Q E R T C
 A A S G I S S L K G G N T V I H R I R G R R K E H
 R R A T D A E F A T I R N D V S D P S L A P F F M
 Q P S R Y S V L R Y Y P G C F P I V A C S L V H
 A A L P I L E S L P P L V T I W L I L P Y R L F S C T

FIG. 3E

8/17

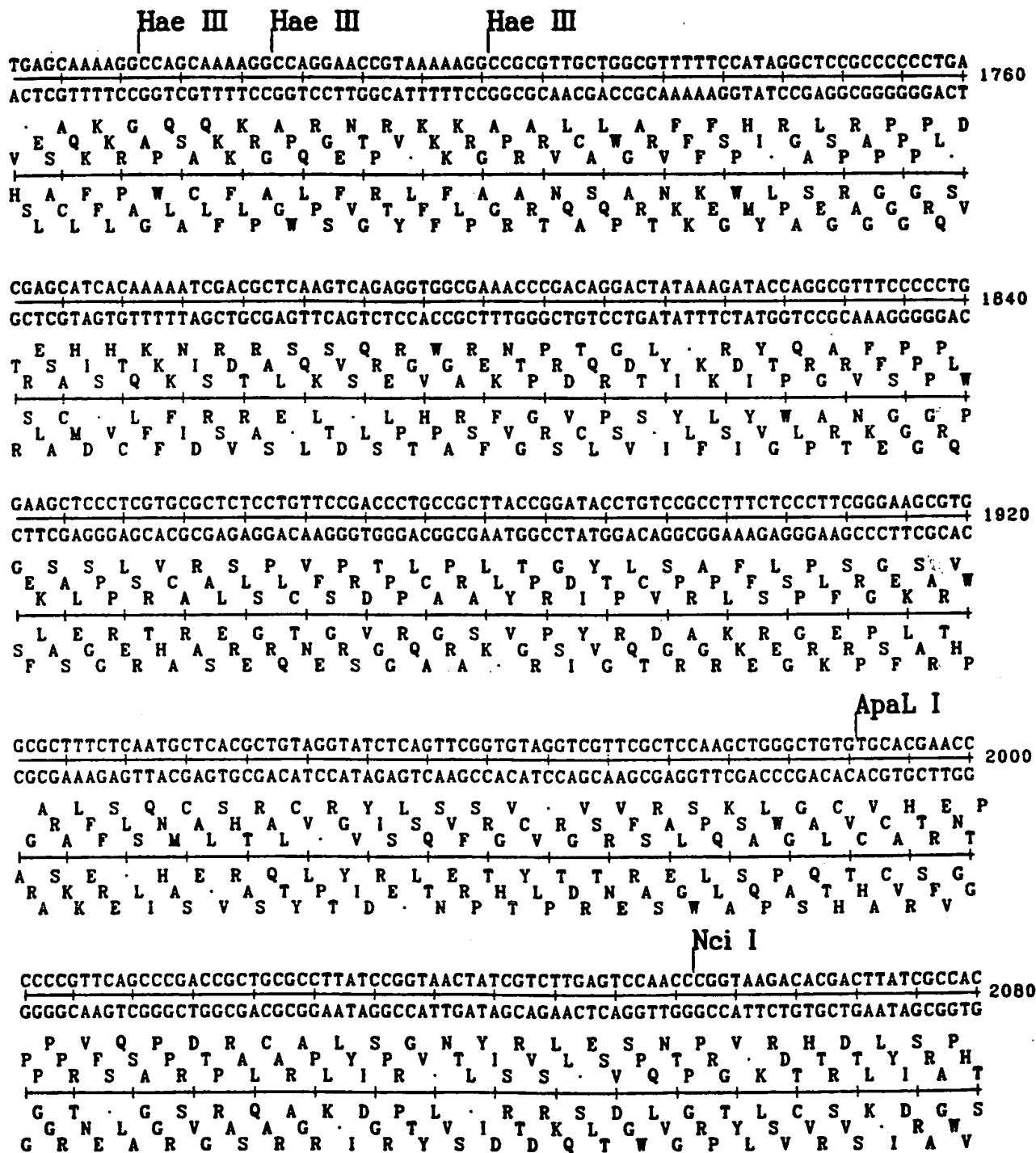


FIG. 3F

9/17

Hae III

TGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAAC 2160
 ACCGTCGTCGGTGACCATTTGCTCTAATCGTCTCGCTCCATACATCCGCCACGATGTCTCAAGAACTTCACCACCGGATTG
 L A A A T G N R I S R A R Y V G G A T E F L K W W P N
 W Q Q P L V T G L A E R G M A V L Q S S S G G L T
 G S S H W Q D Q S E V C R R C Y R V L E V V A
 A A A V P L L I L L A L Y T P P A V S N K F H H G L
 Q C C G S T V P N A S R P I Y A T S C L E Q L P P R V
 P L L W Q Y C S C L S T H L R H L T R S T T A S

TACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAGAGAGTTGGTAGCTC 2240
 ATGCCGATGTGATCTTCCTGTCATAAACCATAGACGCGAGACGACTTCGGTCAATGGAAGCCTTTTCTCAACCATCGAG
 Y G Y T R R T V P G I C A L L K P V T F G K R V G S S
 L R L H K D S I W Y L R S A E A S Y L R K K S W L
 P V L L V T N P I Q A R S F G T V K P F L T P L E
 V A V S S P C Y K T D A S Q Q L W N G E S F S N T A R
 R S C F S L I Q Y R R E A S A L R R F F L Q Y S

TTGATCCGGCAAAACAAACCACCGCTGGTAGCGGTGTTTTTTCTTTGCAAGCAGCAGATTACGGCGAGAAAAAAGGAT 2320
 AACTAGGCCGTTTTGTTGGTGGCGACCATCGCCACCAAAAAACAACGTTTCGTCTAATGCCGCTTTTTTTTCTTA
 S G K Q T T A G S G G F F V C K Q Q I T R R K K G
 L D P A N K P P L V A V V F L F A S S R L R A E K K D
 L I R Q T N H R W R W F F C L Q A A D Y A Q K K R I
 Q D P L C V V A P L P P K K T Q L C C I V R L F F P D
 S G A F L G G S T A T T K K N A L L L N R A S F F S
 K I R C V F W R Q Y R H N K Q K C A A S A C F F L I

BspH I

CTCAAGAAGATCCTTTGATCTTTTCTACGGGTCTGACGCTCAGTGAACGAAAACCTCAGTTAAGGGATTTTGGTCATG 2400
 GAGTTCCTTAGGAACTAGAAAAGATGCCCCAGACTGCGAGTCACCTTGCTTTTGAGTGAATTCCTAAAACAGTAC
 S Q E D P L I F S T G S D A Q W N E N S R G I L V M
 L K K I L S F L R G L T L S G T K T H V K G F W S
 S R R S F D L F Y G V R S V E R K L T L R D F G H
 S S G K I K E V P D S A H F S F E R P I K T M
 R L F I R Q Q K R R P R V S L P V F V T L P N Q D H
 E L L D K S R K P T Q R E T S R F S V N L S K P S

Dra I

Dra I

AGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAATAAATGAAGTTTAAATCAATCTAAAGTATATATGAGTA 2480
 TCTAATAGTTTTTCTAGAAAGTGATCTAGGAAAATTTAATTTTACTTCAAAATTTAGTTAGATTTTCATATATACTCAT
 R L S K R I F T I L L N K S P K S I S I Y E
 D Y Q K G S S P R S F I K N E V L N Q S K V Y M S
 E I I K K D L H L D P F K L K M K F I N L K Y I V
 L N D F L I K V I R K F F H L K L D I L I Y S Y L
 S I I L F S R R S G K L N F I F N I L R F Y I H T

FIG. 3G

10/17

AACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTTG 2580
TTGAACCAGACTGTCAATGGTTACGAATTAGTCACTCCGTGGATAGAGTCGCTAGACAGATAAAGCAAGTAGGTATCAAC
T W S D S Y Q C L I S E A P I S A I C L F R S S I V
K L G L T V T P N A L S V R H L S Q R S V Y F V H P L
N L V Q L P M L N Q G T Y L S D L S I S F I H S C
V Q D S L W H K I L S A G I E A I Q R N R E D M T A
F K T Q C N G I S L D T L P V R L S R D I E N M W L Q

Hae III

CCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGAC 2640
GGACTGAGGGGCAGCACATCTATTGATGCTATGCCCTCCGAATGGTAGACCGGGGTCACGACGTTACTATGGCGCTCTG
A L P V V I T T I R E G L P S G P S A A M I P R D
P D S P S C R L R Y G R A L Y H L A P V L Q N D T A R
L T P R R V D N Y D T G G L T I W P Q C C N D T A R
Q S G T T Y I V V I R S P K G D P G L A A I I G R S V
G S E G R D H L Y S R Y P L A V W R A G T H S C H Y R S V
R V G R R T S L S V P P S V M Q G W H Q L S V A L G

Bgl I

Hae III

Ava II

CCACGCTCACCGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCTGCAACTTT 2720
GGTGCGAGTGGCCGAGGTCTAAATAGTCGTTATTTGGTCGGTCCGGCTTCCCGGCTCGCGTCTTCACCAGGACGTTGAAA
P R S P A P D L S A I N Q P A G R A E R R S G P A T L
H A H R L Q I Y Q Q I T S Q A P E G R A Q K W S C N F
P T L T G S R F I S N K P A S R K G R A Q K W S C N F
G R E G A G S K D A I F W G A P L A S R L L P G A V K
W A S V P E L N I L C Y L G A L R F P R A C F H D Q L K
V S V P E L N I L C Y L G A L R F P R A C F H D Q L K

AseI

Nci I

Fsp I

ATCCGCTCCATCCAGTCTATTAATTGTTGCCGGAAGCTAGAGTAAGTAGTTCCGCCAGTTAATAGTTTGGCGCAACGTTG 2800
TAGGCGGAGGTAGGTACAGATAATTAACAACGGCCCTTCGATCTCATTTCATCAAGCGGTCAATTATCAAACGCGTTGCAAC
S A S I Q S I N C R E A R V S S P V N S L R N V
Y P P P S S L L I V A G K L E V V R Q L I V C A T L
I R L H P V Y L L P G S S K F A S F A Q R C
D A E M W D I L Q Q R S A L T L L E G T L L K R L T T
I G G G D L R N I T A P F S S Y L T R W N I T Q A V N Q
I R R W G T N N G P L L L Y N A L Y N A C R Q

TTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCGTTTGGTATGGCTTCATTACGCTCCGGTTCCTCAACGATCAAGG 2880
AACGGTAACGATGTCCGTAGCACCACAGTGCGAGCAGCAAACCATACCGAAGTAAGTCGAGGCCAAGGGTTGCTAGTTCC
V A I A T G I V V S R S S F G M A S F S S G S Q R S R
L P L L Q A S W C H A R R L V W L H S A P V P N D Q G
C H C Y R A H S G V T L V V W Y G F I Q L R F P T I K
A M A V P M T T D R E D N P I A E N L E P E W R D L
N G N S C A D H T A R R K T H S E A G T G L S I L P
Q W Q L C R P T V S T T Q Y P K M S R N G L V I L A

FIG. 3H

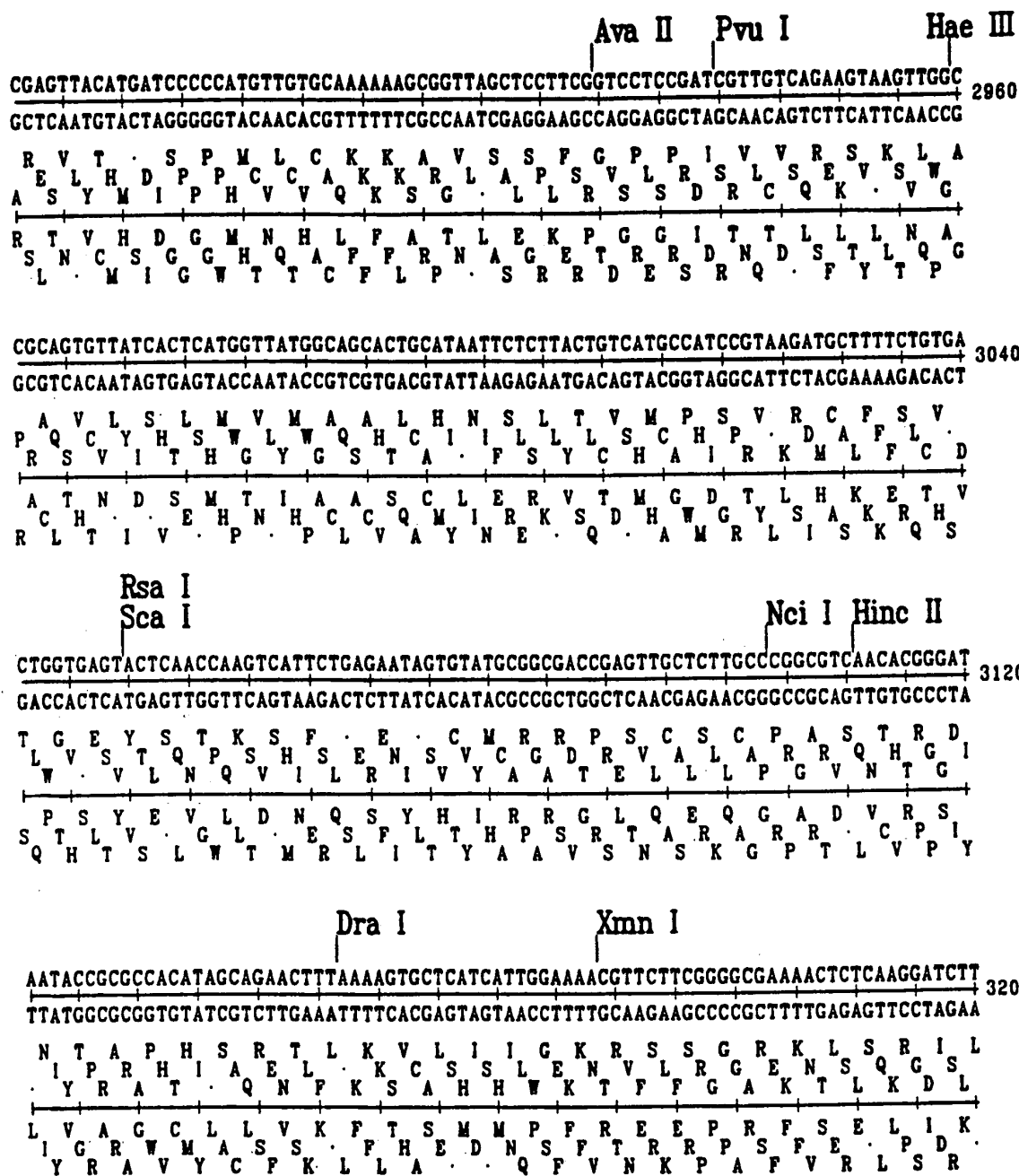


FIG. 31

12/17

ApaI I

ACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACATGATCTTCAGCATCTTTTACTTTACACGCGTTT 3280
TGGCGACAACCTCTAGGTCAAGCTACATTGGGTGAGCACGTGGGTTGACTAGAAAGTCGTAGAAAATGAAAGTGGTCGCAAA
P L L R S S S M P T R A P N S S A S F T F T S V
Y R C D P V R C N P L V H P T D L Q H L L L S P A F
T A V E I Q F D V T H S C T Q L I F S I F Y F H Q R F
G S N L D L E I Y G V R A G L Q D E A D K V K V L T E
V R Q Q S I G T R H L G S T C G V S R C R K S E G A N
V A T S I W N S T V W E H V W S I K L M K K W R K

CTGGGTGAGCAAAACAGGAAGGCAAAATGCCGCAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTC 3380
GACCCACTCGTTTTGTCTTCCGTTTTACGGCGTTTTTCCCTTATTCGCGTGTGCCTTTACAACCTATGAGTATGAG
S G A K T G R Q N A A K K G I R A T R K C I L I L
L G E Q K Q E G K M P Q K K G E G R D H G N V E Y S Y S
W V S K N R K A K C R K K G N K G D T E M L N T H T
P H A F V P L C F A A F F P I L A V R F H Q I S M S
P P S C F C S P L I A C F L S Y P R C P F T S Y E Y E
Q T L L F L F A F H R L F P F L P S V S I N S F V V R

Hinc II

Spe I Asel

TTCCTTTTTCAATATTATTGAAGCATTATCAGGGTTATTGTCTCATGCGCGTTGACATTGATTATTGACTAGTTATTAA 3440
AAGGAAAAAGTTATAATAACTTCGTAATAGTCCCAATAACAGAGTACGCGCAACTGTAACCTAATAACTGATCAATAATT
F L F Q Y Y S I Y Q G Y C L M R V D I D Y L V I N
L P F S I L L K H L S G L Y S H A R H L L T S Y
K R K Y Q L M P Q R M R T S M S Q S T I L
E K K L I I S A N I L T I T E H A N V N I I S N I
G K E I N N F C K D P N N D A R Q C Q N N V L N I

Hae III
Bgl I

TAGTAATCAATTACGGGGTCATTAGTTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCC 3520
ATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATATACCTCAAGGCGCAATGTATTGAATGCCATTTACGGGGCGG
S N Q L R G H F I A H I W S S A L H N L R M A R
I V I N Y G V I S S P I Y G V P R Y I T Y G K W P A
S I T G S L V H S P Y M E F R V T L T V N G P P
L L N R P N M A W I H L E A N C L K R Y I A R R
T I L P T M L E Y G M Y P T G R M V P L H G A
Y Y D I V P D N T L G Y I S N R T V Y S V T F P G A

FIG. 3J

Aat II

TGGCTGACCGCCCAACGACCCCGCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCC 3600
 ACCGACTGGCGGGTTGCTGGGGCGGGTAAGTGCAGTTATTACTGCATACAAGGGTATCATTGCGGTTATCCCTGAAAGG
 L A D R P T T P A H · R Q · · R M F P · · R Q · G L S
 W L T A Q R P P P I D V N N D V C S H S N A N R D F P
 G · P P N D P R P L T S I M T Y V P I V T P I G T F
 A S R G V V G A W Q R · Y H R I N G Y Y R W Y P S E
 Q S V A W R G G M S T L L S T H E W L L A L L S K G
 P Q C G L S G R G N V D I I V Y T G M T V G I P V K W

Aat II

Bgl I

Rsa I

Nde I

Rsa I

ATTGACGTCAATGGGTGGACTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCC 3680
 TAACTGCAGTTACCCACCTGATAAATGCCATTGACGGGTGAACCGTCATGTAGTTCACATAGTATACGGTTCATGCGGG
 I D V N G W T I Y G K L P T W Q Y I K C I I C Q V R P
 L T S M G L F T V N C P L G S T S S V S Y A K Y A
 H · R Q W V D Y L R · T A H L A V H Q V Y H M P S T P
 M S T L P H V I · P L S G V Q C Y M L H I M H W T R G
 N V D I P P S N V T F Q G S P L V D L T D Y A L Y A G
 Q R · H T S · K R Y V A W K A T C · T Y · I G L V G

Hae III

Aat II

Bgl I

Rsa I

CCTATTGACGTCAATGACGGTAAATGCCCCGCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCA 3760
 GGATAACTGCAGTTACTGCCATTTACCGGGCGGACCGTAATACGGGTCTACTGGAATACCCTGAAAGGATGAACCGT
 L L T S M T V N G P P G I M P S T · P Y G T F L L G
 P Y R Q · R · M A R L A L C P V H D L M G L S Y L A
 P I D V N D G K W P A W H Y A Q Y M T L W D P P T W Q
 R N V D I V T F P G G P M I G L V H G · P V K R S P L
 I Q R · H R Y I A R R A N H G T C S R I P S E · K A
 G I S T L S P L H G A Q C · A W Y M V K H S K G V Q C

BsaA I

Nco I

Rsa I

SnaB I

Sty I

Rsa I

GTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTG 3840
 CATGTAGATGCATAATCAGTAGCGATAATGGTACCACTACGCCAAAACCGTCATGTAGTTACCGGCACCTATGCCAAAC
 S T S T Y · S S L L P W · C G F G S T S M G V D S G L
 V H L R I S H R Y Y H G D A V L A V H Q W A W I A V
 Y I Y V L V I A I T M V M R F W Q Y I N G R G · R F
 V D V Y · D D S N G H H H P K P L V D I P T S L P K
 T C R R I L · R · W P S A T K A T C · H A H I A T K
 Y M · T N T M A I V M T I R N Q C Y M L P R P Y R N S

FIG. 3K

14/17

Aat II

ACTCACGGGGATTTCCTCAAGTCTCCACCCCATTTGACGTCAATGGGAGTTTGTGTTTGGCACCAAAATCAACGGGACTTTCCA 3920
TGAGTGCCCCCTAAAGGTTTCAGAGCTGGGGTAACTGCAGTTACCCTCAAACAAAACCGTGGTTTTAGTTGCCCTGAAAGGT
T H G D P Q V S T P L T S M G V C F G T K I N G T F Q
L T G I S K S L P P H R Q W E F V L A P K S T G L S P
D S R G F P S L H P I D V N G S L F W H Q N Q R D P P
V P S K W T E V G N V D I P T Q K P V L I L P V K W
S V P P I E L D G G W Q R H S N T K A G F D V P S E L
E R P N G L R W G M S T L P L K N Q C W F R S K G

Rsa I

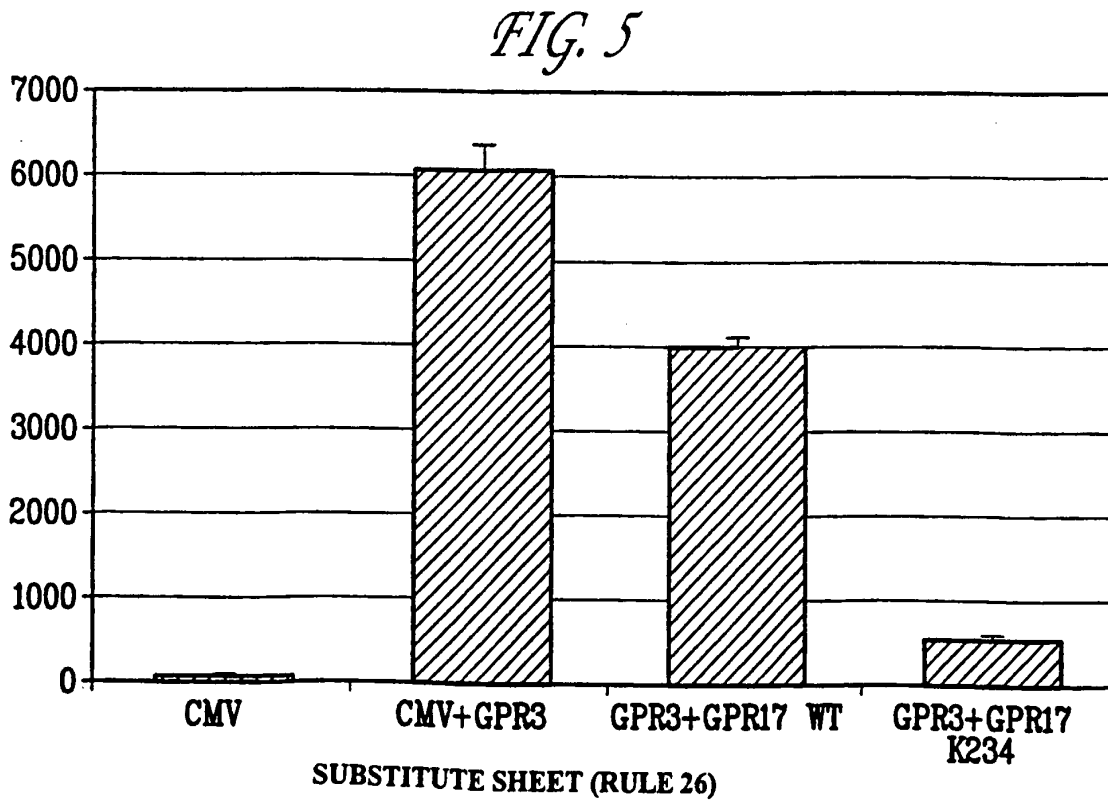
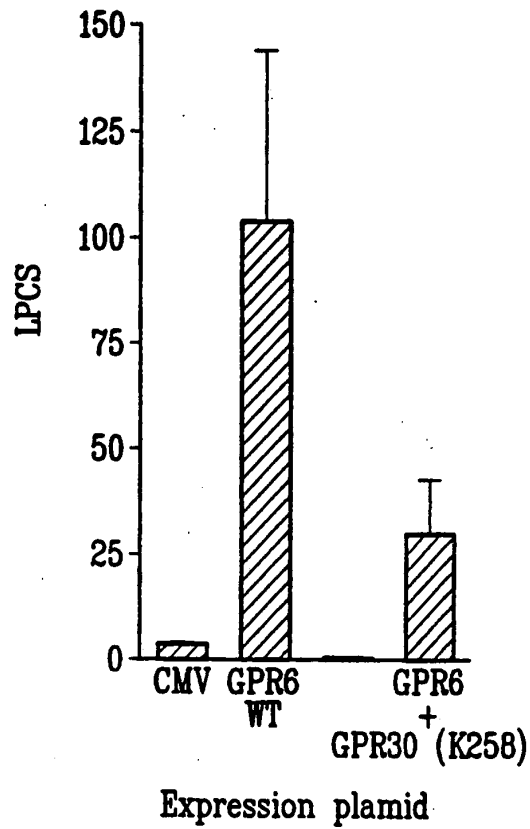
Sac I

AAATGTCGTAACAACCTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCT 4000
TTTACAGCATTGTTGAGGCGGGTAACTGCGTTTACCGCCATCCGCACATGCCACCCTCCAGATATATTCGTCTCGAGA
N V V T T P P H R K W A V G V Y G G R S I A E L
K M S Q L R P I D A N G R A C T V G G L Y K Q S S
K C R N N S A P L T Q M G G R R V R W E V Y I S R A L
F T T V V G G W Q R L H A T P T Y P P L D I Y A S S E
F I D Y C S R G M S A F P R Y A H V T P P R Y L C L E
F H R L L E A G N V C I P P L R T R H S T I L L A R

Rsa I

CTGGCTAACTAGAGAACCCACTGCTTAACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCC 4069
GACGGATTGATCTCTTGGGTGACGAATTGACCGAATAGCTTTAATTATGCTGAGTGATATCCCTCTGGG
S G L E N P L L N W L I E I N T T H Y R E T
L A N R E P T H C L T G L S K L I R L T I G R P
W L T R E P T A L A Y R N Y D S L G D P
P S S F G S S L Q S I S I L V V L S V W
R A L L V W Q K V P K D F N I R S V I P L G W
Q S V L S G V A S A R F Y S E S Y P S G

FIG. 3L



16/17

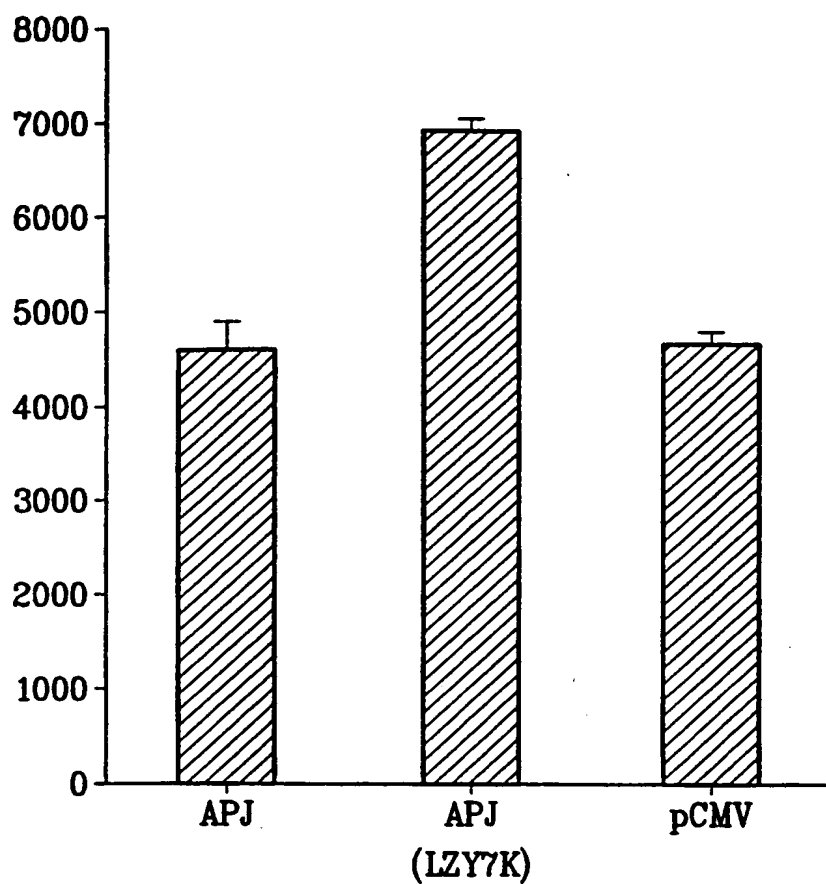
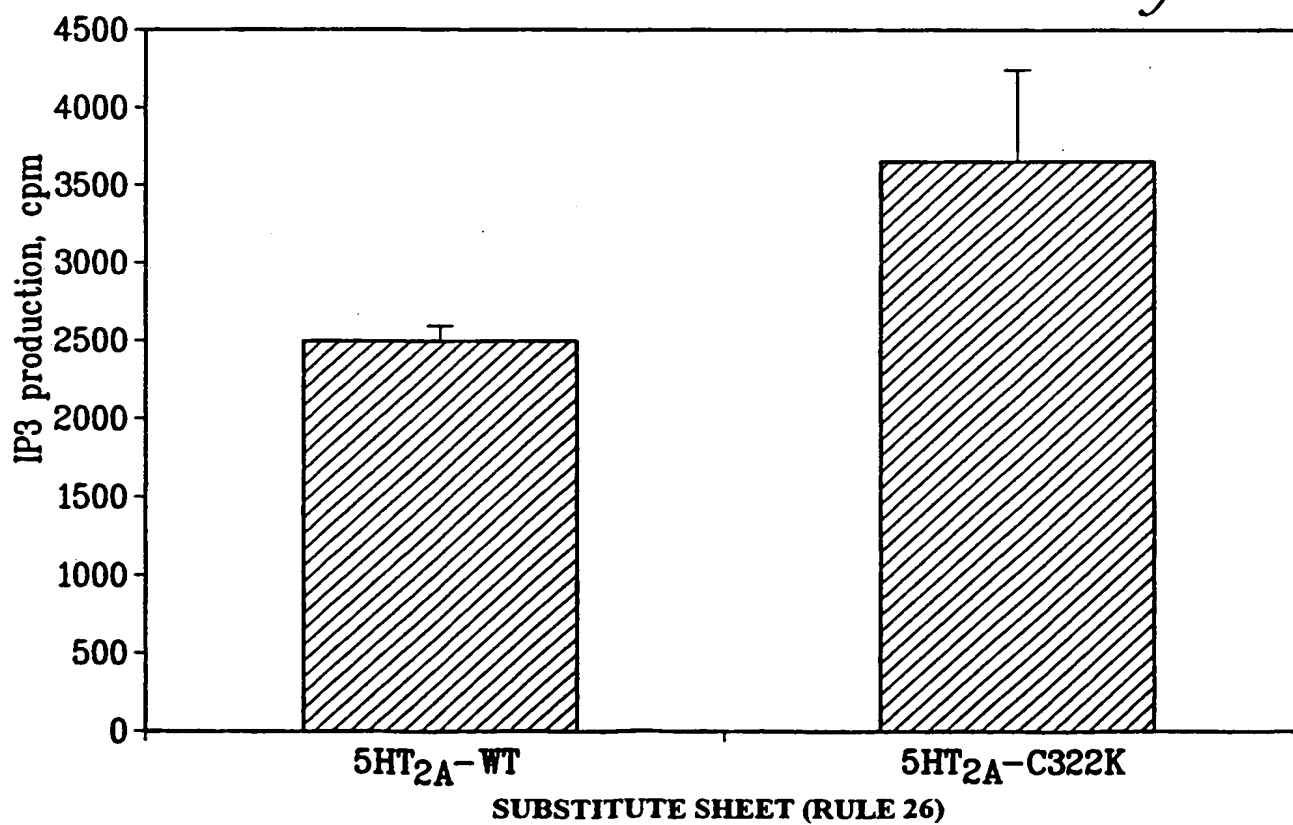
*FIG. 6**FIG. 7*



FIG. 8A

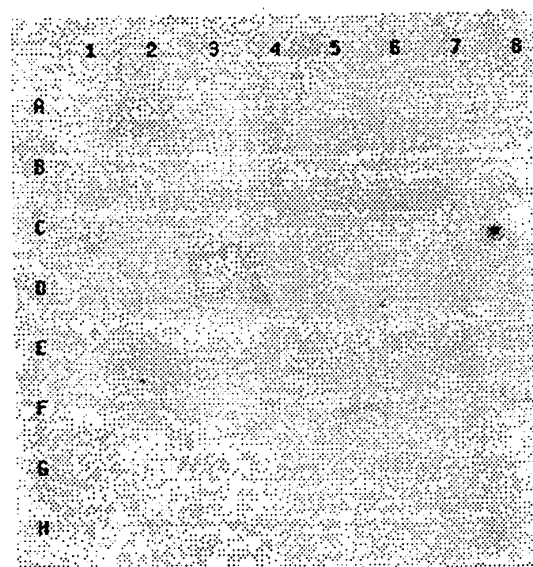


FIG. 8B

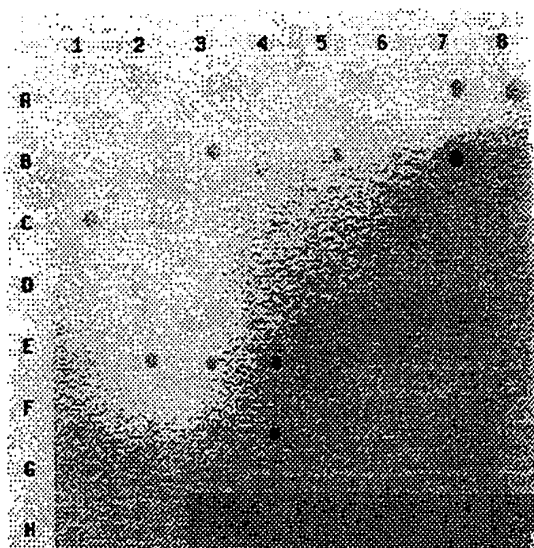


FIG. 8C

1

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- 5 (i) APPLICANT: Behan, Dominic P.
Chalmers, Derek T.
Liaw, Chen W.
- (ii) TITLE OF INVENTION: Non-Endogenous, Constitutively
Activated Human G Protein-Coupled
Orphan Receptors
- 10 (iii) NUMBER OF SEQUENCES: 280
- (iv) CORRESPONDENCE ADDRESS:
15 (A) ADDRESSEE: Arena Pharmaceuticals, Inc.
(B) STREET: 6166 Nancy Ridge Drive
(C) CITY: San Diego
(D) STATE: CA
(E) COUNTRY: USA
(F) ZIP: 92122
- (v) COMPUTER READABLE FORM:
20 (A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
25 (A) APPLICATION NUMBER: US
(B) FILING DATE:
(C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Burgoon, Richard P.
(B) REGISTRATION NUMBER: 34,787
- 30 (ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (619)453-7200
(B) TELEFAX: (619)453-7210

(2) INFORMATION FOR SEQ ID NO:1:

- 35 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1068 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- 40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATGGAAGATT TGGAGGAAAC ATTATTTGAA GAATTTGAAA ACTATTCCTA TGACCTAGAC

60

TATTACTCTC TGGAGTCTGA TTTGGAGGAG AAAGTCCAGC TGGGAGTTGT TCACTGGGTC 120
 TCCCTGGTGT TATATTGTTT GGCTTTTGT CTGGGAATTC CAGGAAATGC CATCGTCATT 180
 TGGTTCACGG GGCTCAAGTG GAAGAAGACA GTCACCACTC TGTGGTTCCT CAATCTAGCC 240
 ATTGCGGATT TCATTTTCT TCTCTTCTG CCCCTGTACA TCTCCTATGT GGCCATGAAT 300
 5 TTCCACTGGC CCTTTGGCAT CTGGCTGTGC AAAGCCAATT CCTTCACTGC CCAGTTGAAC 360
 ATGTTTGCCA GTGTTTTTTT CCTGACAGTG ATCAGCCTGG ACCACTATAT CCACTTGATC 420
 CATCCTGTCT TATCTCATCG GCATCGAACC CTCAAGAACT CTCTGATTGT CATTATATTC 480
 ATCTGGCTTT TGGCTTCTCT AATTGGCGGT CCTGCCCTGT ACTCCGGGA CACTGTGGAG 540
 TTCAATAATC ATACTCTTTG CTATAACAAT TTTCAGAAGC ATGATCCTGA CCTCACTTGT 600
 10 ATCAGGCACC ATGTTCTGAC TTGGGTGAAA TTTATCATTG GCTATCTCTT CCCTTTGCTA 660
 ACAATGAGTA TTTGCTACTT GTGTCTCATC TTCAAGGTGA AGAAGCGAAC AGTCCTGATC 720
 TCCAGTAGGC ATTTCTGGAC AATTCTGGTT GTGGTTGTGG CCTTGTGGT TTGCTGGACT 780
 CCTTATCACC TGTTTAGCAT TTGGGAGCTC ACCATTACAC ACAATAGCTA TTCCCACCAT 840
 GTGATGCAGG CTGGAATCCC CCTCTCCACT GGTGTGGCAT TCCTCAATAG TTGCTTGAAC 900
 15 CCCATCCTTT ATGTCCTAAT TAGTAAGAAG TTCCAAGCTC GCTTCCGGTC CTCAGTTGCT 960
 GAGATACTCA AGTACACACT GTGGGAAGTC AGCTGTTCTG GCACAGTGAG TGAACAGCTC 1020
 AGGAACTCAG AAACCAAGAA TCTGTGTCTC CTGGAAACAG CTCAATAA 1068

(3) INFORMATION FOR SEQ ID NO:2:

20 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 355 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Glu Asp Leu Glu Glu Thr Leu Phe Glu Glu Phe Glu Asn Tyr Ser
 1 5 10 15

Tyr Asp Leu Asp Tyr Tyr Ser Leu Glu Ser Asp Leu Glu Glu Lys Val
 20 25 30

30 Gln Leu Gly Val Val His Trp Val Ser Leu Val Leu Tyr Cys Leu Ala
 35 40 45

Phe Val Leu Gly Ile Pro Gly Asn Ala Ile Val Ile Trp Phe Thr Gly
 50 55 60
 Leu Lys Trp Lys Lys Thr Val Thr Thr Leu Trp Phe Leu Asn Leu Ala
 65 70 75 80
 5 Ile Ala Asp Phe Ile Phe Leu Leu Phe Leu Pro Leu Tyr Ile Ser Tyr
 85 90 95
 Val Ala Met Asn Phe His Trp Pro Phe Gly Ile Trp Leu Cys Lys Ala
 100 105 110
 10 Asn Ser Phe Thr Ala Gln Leu Asn Met Phe Ala Ser Val Phe Phe Leu
 115 120 125
 Thr Val Ile Ser Leu Asp His Tyr Ile His Leu Ile His Pro Val Leu
 130 135 140
 Ser His Arg His Arg Thr Leu Lys Asn Ser Leu Ile Val Ile Ile Phe
 145 150 155 160
 15 Ile Trp Leu Leu Ala Ser Leu Ile Gly Gly Pro Ala Leu Tyr Phe Arg
 165 170 175
 Asp Thr Val Glu Phe Asn Asn His Thr Leu Cys Tyr Asn Asn Phe Gln
 180 185 190
 20 Lys His Asp Pro Asp Leu Thr Leu Ile Arg His His Val Leu Thr Trp
 195 200 205
 Val Lys Phe Ile Ile Gly Tyr Leu Phe Pro Leu Leu Thr Met Ser Ile
 210 215 220
 Cys Tyr Leu Cys Leu Ile Phe Lys Val Lys Lys Arg Thr Val Leu Ile
 225 230 235 240
 25 Ser Ser Arg His Phe Trp Thr Ile Leu Val Val Val Val Ala Phe Val
 245 250 255
 Val Cys Trp Thr Pro Tyr His Leu Phe Ser Ile Trp Glu Leu Thr Ile
 260 265 270
 30 His His Asn Ser Tyr Ser His His Val Met Gln Ala Gly Ile Pro Leu
 275 280 285
 Ser Thr Gly Leu Ala Phe Leu Asn Ser Cys Leu Asn Pro Ile Leu Tyr
 290 295 300
 Val Leu Ile Ser Lys Lys Phe Gln Ala Arg Phe Arg Ser Ser Val Ala
 305 310 315 320
 35 Glu Ile Leu Lys Tyr Thr Leu Trp Glu Val Ser Cys Ser Gly Thr Val
 325 330 335
 Ser Glu Gln Leu Arg Asn Ser Glu Thr Lys Asn Leu Cys Leu Leu Glu

340

345

350

Thr Ala Gln
355

(4) INFORMATION FOR SEQ ID NO:3:

- 5 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1089 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

- 10 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATGGGCAACC ACACGTGGGA GGGCTGCCAC GTGGACTCGC GCGTGGACCA CCTCTTTCCG 60
 CCATCCCTCT ACATCTTTGT CATCGGCGTG GGGCTGCCCC CCAACTGCCT GGCTCTGTGG 120
 GCGGCCTACC GCCAGGTGCA ACAGCGCAAC GAGCTGGGCG TCTACCTGAT GAACCTCAGC 180
 15 ATCGCCGACC TGCTGTACAT CTGCACGCTG CCGCTGTGGG TGGACTACTT CCTGCACCAC 240
 GACAACTGGA TCCACGGCCC CGGGTCCTGC AAGCTCTTTG GGTTCATCTT CTACACCAAT 300
 ATCTACATCA GCATCGCCTT CCTGTGCTGC ATCTCGGTGG ACCGCTACCT GGCTGTGGCC 360
 CACCCACTCC GCTTCGCCCC CCTGCGCCGC GTCAAGACCG CCGTGGCCGT GAGCTCCGTG 420
 GTCTGGGCCA CGGAGCTGGG CGCCAACTCG GCGCCCCTGT TCCATGACGA GCTCTTCCGA 480
 20 GACCGCTACA ACCACACCTT CTGCTTTGAG AAGTTCCCCA TGGGAAGGCTG GGTGGCCTGG 540
 ATGAACCTCT ATCGGGTGTT CGTGGGCTTC CTCTTCCCGT GGGCGCTCAT GCTGCTGTCTG 600
 TACCGGGGCA TCCTGCGGGC CGTGCGGGGC AGCGTGTTCA CCGAGCGCCA GGAGAAGGCC 660
 AAGATCAAGC GGCTGGCCCT CAGCCTCATC GCCATCGTGC TGGTCTGCTT TGCGCCCTAT 720
 CACGTGCTCT TGCTGTCCCG CAGCGCCATC TACCTGGGCC GCCCCTGGGA CTGCGGCTTC 780
 25 GAGGAGCGCG TCTTTTCTGC ATACCACAGC TCACTGGCTT TCACCAGCCT CAACTGTGTG 840
 GCGGACCCCA TCCTCTACTG CCTGGTCAAC GAGGGCGCCC GCAGCGATGT GGCCAAGGCC 900
 CTGCACAACC TGCTCCGCTT TCTGGCCAGC GACAAGCCCC AGGAGATGGC CAATGCCTCG 960
 CTCACCCTGG AGACCCCACT CACCTCCAAG AGGAACAGCA CAGCCAAAGC CATGACTGGC 1020
 AGCTGGGCGG CCACTCCGCC TTCCCAGGGG GACCAGGTGC AGCTGAAGAT GCTGCCGCCA 1080
 30 GCACAATGA 1089

(5) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 362 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

10 Met Gly Asn His Thr Trp Glu Gly Cys His Val Asp Ser Arg Val Asp
 1 5 10 15
 His Leu Phe Pro Pro Ser Leu Tyr Ile Phe Val Ile Gly Val Gly Leu
 20 25 30
 Pro Thr Asn Cys Leu Ala Leu Trp Ala Ala Tyr Arg Gln Val Gln Gln
 35 40 45
 15 Arg Asn Glu Leu Gly Val Tyr Leu Met Asn Leu Ser Ile Ala Asp Leu
 50 55 60
 Leu Tyr Ile Cys Thr Leu Pro Leu Trp Val Asp Tyr Phe Leu His His
 65 70 75 80
 20 Asp Asn Trp Ile His Gly Pro Gly Ser Cys Lys Leu Phe Gly Phe Ile
 85 90 95
 Phe Tyr Thr Asn Ile Tyr Ile Ser Ile Ala Phe Leu Cys Cys Ile Ser
 100 105 110
 Val Asp Arg Tyr Leu Ala Val Ala His Pro Leu Arg Phe Ala Arg Leu
 115 120 125
 25 Arg Arg Val Lys Thr Ala Val Ala Val Ser Ser Val Val Trp Ala Thr
 130 135 140
 Glu Leu Gly Ala Asn Ser Ala Pro Leu Phe His Asp Glu Leu Phe Arg
 145 150 155 160
 30 Asp Arg Tyr Asn His Thr Phe Cys Phe Glu Lys Phe Pro Met Glu Gly
 165 170 175
 Trp Val Ala Trp Met Asn Leu Tyr Arg Val Phe Val Gly Phe Leu Phe
 180 185 190
 Pro Trp Ala Leu Met Leu Leu Ser Tyr Arg Gly Ile Leu Arg Ala Val
 195 200 205
 35 Arg Gly Ser Val Ser Thr Glu Arg Gln Glu Lys Ala Lys Ile Lys Arg
 210 215 220
 Leu Ala Leu Ser Leu Ile Ala Ile Val Leu Val Cys Phe Ala Pro Tyr

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1002 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ATGGAGTCCT CAGGCAACCC AGAGAGCACC ACCTTTTTTTT ACTATGACCT TCAGAGCCAG 60
 CCGTGTGAGA ACCAGGCCTG GGTCTTTGCT ACCCTCGCCA CCACTGTCCT GTACTGCCTG 120
 10 GTGTTTCTCC TCAGCCTAGT GGGCAACAGC CTGGTCCTGT GGGTCCTGGT GAAGTATGAG 180
 AGCCTGGAGT CCCTCACCAA CATCTTCATC CTCAACCTGT GCCTCTCAGA CCTGGTGTTT 240
 GCCTGCTTGT TGCCTGTGTG GATCTCCCCA TACCACTGGG GCTGGGTGCT GGGAGACTTC 300
 CTCTGCAAAC TCCTCAATAT GATCTTCTCC ATCAGCCTCT ACAGCAGCAT CTTCTTCCTG 360
 ACCATCATGA CCATCCACCG CTACCTGTCTG GTAGTGAGCC CCCTCTCCAC CCTGCGCGTC 420
 15 CCCACCCTCC GCTGCCGGGT GCTGGTGACC ATGGCTGTGT GGGTAGCCAG CATCCTGTCC 480
 TCCATCCTCG ACACCATCTT CCACAAGGTG CTTTCTTCGG GCTGTGATTA TTCCGAACTC 540
 ACGTGGTACC TCACCTCCGT CTACCAGCAC AACCTCTTCT TCCTGCTGTC CCTGGGGATT 600
 ATCCTGTTCT GCTACGTGGA GATCCTCAGG ACCCTGTTCC GCTCACGCTC CAAGCGGCGC 660
 CACCGCACGG TCAAGCTCAT CTTGCCCATC GTGGTGGCCT ACTTCCTCAG CTGGGGTCCC 720
 20 TACAACTTCA CCCTGTTTCT GCAGACGCTG TTTCGGACCC AGATCATCCG GAGCTGCGAG 780
 GCCAAACAGC AGCTAGAATA CGCCCTGCTC ATCTGCCGCA ACCTCGCCTT CTCCCCTGTC 840
 TGCTTTAACC CGGTGCTCTA TGTCTTCGTG GGGGTCAAGT TCCGCACACA CCTGAAACAT 900
 GTTCTCCGGC AGTTCTGGTT CTGCCGGCTG CAGGCACCCA GCCCAGCCTC GATCCCCCAC 960
 TCCCCTGGTG CCTTCGCCTA TGAGGGCGCC TCCTTCTACT GA 1002

25 (9) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 333 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

	Met	Glu	Ser	Ser	Gly	Asn	Pro	Glu	Ser	Thr	Thr	Phe	Phe	Tyr	Tyr	Asp	
	1				5					10					15		
	Leu	Gln	Ser	Gln	Pro	Cys	Glu	Asn	Gln	Ala	Trp	Val	Phe	Ala	Thr	Leu	
				20					25					30			
5	Ala	Thr	Thr	Val	Leu	Tyr	Cys	Leu	Val	Phe	Leu	Leu	Ser	Leu	Val	Gly	
			35					40					45				
	Asn	Ser	Leu	Val	Leu	Trp	Val	Leu	Val	Lys	Tyr	Glu	Ser	Leu	Glu	Ser	
		50					55					60					
10	Leu	Thr	Asn	Ile	Phe	Ile	Leu	Asn	Leu	Cys	Leu	Ser	Asp	Leu	Val	Phe	
	65					70					75					80	
	Ala	Cys	Leu	Leu	Pro	Val	Trp	Ile	Ser	Pro	Tyr	His	Trp	Gly	Trp	Val	
					85					90					95		
	Leu	Gly	Asp	Phe	Leu	Cys	Lys	Leu	Leu	Asn	Met	Ile	Phe	Ser	Ile	Ser	
				100					105					110			
15	Leu	Tyr	Ser	Ser	Ile	Phe	Phe	Leu	Thr	Ile	Met	Thr	Ile	His	Arg	Tyr	
			115					120					125				
	Leu	Ser	Val	Val	Ser	Pro	Leu	Ser	Thr	Leu	Arg	Val	Pro	Thr	Leu	Arg	
		130					135					140					
20	Cys	Arg	Val	Leu	Val	Thr	Met	Ala	Val	Trp	Val	Ala	Ser	Ile	Leu	Ser	
	145					150					155					160	
	Ser	Ile	Leu	Asp	Thr	Ile	Phe	His	Lys	Val	Leu	Ser	Ser	Gly	Cys	Asp	
				165						170					175		
	Tyr	Ser	Glu	Leu	Thr	Trp	Tyr	Leu	Thr	Ser	Val	Tyr	Gln	His	Asn	Leu	
			180					185						190			
25	Phe	Phe	Leu	Leu	Ser	Leu	Gly	Ile	Ile	Leu	Phe	Cys	Tyr	Val	Glu	Ile	
			195				200					205					
	Leu	Arg	Thr	Leu	Phe	Arg	Ser	Arg	Ser	Lys	Arg	Arg	His	Arg	Thr	Val	
		210					215					220					
30	Lys	Leu	Ile	Phe	Ala	Ile	Val	Val	Ala	Tyr	Phe	Leu	Ser	Trp	Gly	Pro	
	225					230					235					240	
	Tyr	Asn	Phe	Thr	Leu	Phe	Leu	Gln	Thr	Leu	Phe	Arg	Thr	Gln	Ile	Ile	
				245					250						255		
	Arg	Ser	Cys	Glu	Ala	Lys	Gln	Gln	Leu	Glu	Tyr	Ala	Leu	Leu	Ile	Cys	
			260					265						270			
35	Arg	Asn	Leu	Ala	Phe	Ser	His	Cys	Cys	Phe	Asn	Pro	Val	Leu	Tyr	Val	
		275						280					285				

Phe Val Gly Val Lys Phe Arg Thr His Leu Lys His Val Leu Arg Gln
290 295 300

Phe Trp Phe Cys Arg Leu Gln Ala Pro Ser Pro Ala Ser Ile Pro His
305 310 315 320

5 Ser Pro Gly Ala Phe Ala Tyr Glu Gly Ala Ser Phe Tyr
325 330

(10) INFORMATION FOR SEQ ID NO:9:

10 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

15 GCAAGCTTGG GGGACGCCAG GTCGCCGGCT 30

(11) INFORMATION FOR SEQ ID NO:10:

20 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GCGGATCCGG ACGCTGGGGG AGTCAGGCTG C 31

25 (12) INFORMATION FOR SEQ ID NO:11:

30 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 987 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

ATGGACAACG CCTCGTTCTC GGAGCCCTGG CCCGCCAACG CATCGGGCCC GGACCCGGCG 60

CTGAGCTGCT CCAACGCGTC GACTCTGGCG CCGTGCCGG CGCCGCTGGC GGTGGCTGTA 120

35 CCAGTTGTCT ACGCGGTGAT CTGCGCCGTG GGTCTGGCGG GCAACTCCGC CGTGCTGTAC 180

GTGTTGCTGC GGGCGCCCCG CATGAAGACC GTCACCAACC TGTTCATCCT CAACCTGGCC 240
 ATCGCCGACG AGCTCTTCAC GCTGGTGCTG CCCATCAACA TCGCCGACTT CCTGCTGCGG 300
 CAGTGGCCCT TCGGGGAGCT CATGTGCAAG CTCATCGTGG CTATCGACCA GTACAACACC 360
 TTCTCCAGCC TCTACTTCCT CACCGTCATG AGCGCCGACC GCTACCTGGT GGTGTTGGCC 420
 5 ACTGCGGAGT CGCGCCGGGT GGCCGGCCGC ACCTACAGCG CCGCGCGCGC GGTGAGCCTG 480
 GCCGTGTGGG GGATCGTCAC ACTCGTCGTG CTGCCCTTCG CAGTCTTCGC CCGGCTAGAC 540
 GACGAGCAGG GCCGGCGCCA GTGCGTGCTA GTCTTTCCGC AGCCCGAGGC CTTCTGGTGG 600
 CGCGCGAGCC GCCTCTACAC GCTCGTGCTG GGCTTCGCCA TCCCCGTGTC CACCATCTGT 660
 GTCTCTATA CCACCCTGCT GTGCCGGCTG CATGCCATGC GGCTGGACAG CCACGCCAAG 720
 10 GCCCTGGAGC GCGCCAAGAA GCGGGTGACC TTCCTGGTGG TGGCAATCCT GGCGGTGTGC 780
 CTCCTCTGCT GGACGCCCTA CCACCTGAGC ACCGTGGTGG CGCTCACCAC CGACCTCCCG 840
 CAGACGCCGC TGGTCATCGC TATCTCCTAC TTCATCACCA GCCTGACGTA CGCCAACAGC 900
 TGCCTCAACC CCTTCTCTA CGCCTTCTG GACGCCAGCT TCCGCAGGAA CCTCCGCCAG 960
 CTGATAACTT GCCGCGCGGC AGCCTGA 987

15 (13) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 328 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

20

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Met Asp Asn Ala Ser Phe Ser Glu Pro Trp Pro Ala Asn Ala Ser Gly
 1 5 10 15
 25 Pro Asp Pro Ala Leu Ser Cys Ser Asn Ala Ser Thr Leu Ala Pro Leu
 20 25 30
 Pro Ala Pro Leu Ala Val Ala Val Pro Val Val Tyr Ala Val Ile Cys
 35 40 45
 30 Ala Val Gly Leu Ala Gly Asn Ser Ala Val Leu Tyr Val Leu Leu Arg
 50 55 60
 Ala Pro Arg Met Lys Thr Val Thr Asn Leu Phe Ile Leu Asn Leu Ala
 65 70 75 80

11

Ile Ala Asp Glu Leu Phe Thr Leu Val Leu Pro Ile Asn Ile Ala Asp
 85 90 95
 Phe Leu Leu Arg Gln Trp Pro Phe Gly Glu Leu Met Cys Lys Leu Ile
 100 105 110
 5 Val Ala Ile Asp Gln Tyr Asn Thr Phe Ser Ser Leu Tyr Phe Leu Thr
 115 120 125
 Val Met Ser Ala Asp Arg Tyr Leu Val Val Leu Ala Thr Ala Glu Ser
 130 135 140
 10 Arg Arg Val Ala Gly Arg Thr Tyr Ser Ala Ala Arg Ala Val Ser Leu
 145 150 155 160
 Ala Val Trp Gly Ile Val Thr Leu Val Val Leu Pro Phe Ala Val Phe
 165 170 175
 Ala Arg Leu Asp Asp Glu Gln Gly Arg Arg Gln Cys Val Leu Val Phe
 180 185 190
 15 Pro Gln Pro Glu Ala Phe Trp Trp Arg Ala Ser Arg Leu Tyr Thr Leu
 195 200 205
 Val Leu Gly Phe Ala Ile Pro Val Ser Thr Ile Cys Val Leu Tyr Thr
 210 215 220
 20 Thr Leu Leu Cys Arg Leu His Ala Met Arg Leu Asp Ser His Ala Lys
 225 230 235 240
 Ala Leu Glu Arg Ala Lys Lys Arg Val Thr Phe Leu Val Val Ala Ile
 245 250 255
 Leu Ala Val Cys Leu Leu Cys Trp Thr Pro Tyr His Leu Ser Thr Val
 260 265 270
 25 Val Ala Leu Thr Thr Asp Leu Pro Gln Thr Pro Leu Val Ile Ala Ile
 275 280 285
 Ser Tyr Phe Ile Thr Ser Leu Thr Tyr Ala Asn Ser Cys Leu Asn Pro
 290 295 300
 30 Phe Leu Tyr Ala Phe Leu Asp Ala Ser Phe Arg Arg Asn Leu Arg Gln
 305 310 315 320
 Leu Ile Thr Cys Arg Ala Ala Ala
 325

(14) INFORMATION FOR SEQ ID NO:13:

- 35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CGGAATTCGT CAACGGTCCC AGCTACAATG

30

(15) INFORMATION FOR SEQ ID NO:14:

- 5 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 31 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: DNA (genomic)
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

ATGGATCCCA GGCCCTTCAG CACCGCAATA T

31

(16) INFORMATION FOR SEQ ID NO:15:

- 15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1002 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

ATGCAGGCCG CTGGGCACCC AGAGCCCCTT GACAGCAGGG GCTCCTTCTC CCTCCCCACG 60
 ATGGGTGCCA ACGTCTCTCA GGACAATGGC ACTGGCCACA ATGCCACCTT CTCCGAGCCA 120
 CTGCCGTTCC TCTATGTGCT CCTGCCCCGC GTGTACTCCG GGATCTGTGC TGTGGGGCTG 180
 ACTGGCAACA CGGCCGTCAT CCTTGTAATC CTAAGGGCGC CCAAGATGAA GACGGTGACC 240
 25 AACGTGTTCA TCCTGAACCT GGCCGTCGCC GACGGGCTCT TCACGCTGGT ACTGCCCCGC 300
 AACATCGCGG AGCACCTGCT GCAGTACTGG CCCTTCGGGG AGCTGCTCTG CAAGCTGGTG 360
 CTGGCCGTCG ACCACTACAA CATCTTCTCC AGCATCTACT TCCTAGCCGT GATGAGCGTG 420
 GACCGATACC TGGTGGTGCT GGCCACCGTG AGGTCCCGCC ACATGCCCTG GCGCACCTAC 480
 CGGGGGGCGA AGGTGCGCAG CCTGTGTGTC TGGCTGGGCG TCACGGTCCT GGTTCGTCCC 540
 30 TTCTTCTCTT TCGCTGGCGT CTACAGCAAC GAGCTGCAGG TCCCAAGCTG TGGGCTGAGC 600
 TTCCCGTGGC CCGAGCGGGT CTGGTTCAAG GCCAGCCGTG TCTACACTTT GGTCTGGGC 660
 TTCGTGCTGC CCGTGTGCAC CATCTGTGTG CTCTACACAG ACCTCCTGCG CAGGCTGCGG 720

GCCGTGCGGC TCCGCTCTGG AGCCAAGGCT CTAGGCAAGG CCAGGCGGAA GGTGACCGTC 780
 CTGGTCCTCG TCGTGCTGGC CGTGTGCCTC CTCTGCTGGA CGCCCTTCCA CCTGGCCTCT 840
 GTCGTGGCCC TGACCACGCA CCTGCCCCAG ACCCCACTGG TCATCAGTAT GTCCTACGTC 900
 ATCACCAGCC TCACGTACGC CAACTCGTGC CTGAACCCCT TCCTCTACGC CTTTCTAGAT 960
 5 GACAACTTCC GGAAGAACTT CCGCAGCATA TTGCGGTGCT GA 1002

(17) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 333 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

15 Met Gln Ala Ala Gly His Pro Glu Pro Leu Asp Ser Arg Gly Ser Phe
 1 5 10 15
 Ser Leu Pro Thr Met Gly Ala Asn Val Ser Gln Asp Asn Gly Thr Gly
 20 25 30
 His Asn Ala Thr Phe Ser Glu Pro Leu Pro Phe Leu Tyr Val Leu Leu
 35 40 45
 20 Pro Ala Val Tyr Ser Gly Ile Cys Ala Val Gly Leu Thr Gly Asn Thr
 50 55 60
 Ala Val Ile Leu Val Ile Leu Arg Ala Pro Lys Met Lys Thr Val Thr
 65 70 75 80
 25 Asn Val Phe Ile Leu Asn Leu Ala Val Ala Asp Gly Leu Phe Thr Leu
 85 90 95
 Val Leu Pro Val Asn Ile Ala Glu His Leu Leu Gln Tyr Trp Pro Phe
 100 105 110
 Gly Glu Leu Leu Cys Lys Leu Val Leu Ala Val Asp His Tyr Asn Ile
 115 120 125
 30 Phe Ser Ser Ile Tyr Phe Leu Ala Val Met Ser Val Asp Arg Tyr Leu
 130 135 140
 Val Val Leu Ala Thr Val Arg Ser Arg His Met Pro Trp Arg Thr Tyr
 145 150 155 160
 35 Arg Gly Ala Lys Val Ala Ser Leu Cys Val Trp Leu Gly Val Thr Val
 165 170 175

Leu Val Leu Pro Phe Phe Ser Phe Ala Gly Val Tyr Ser Asn Glu Leu
 180 185 190
 Gln Val Pro Ser Cys Gly Leu Ser Phe Pro Trp Pro Glu Arg Val Trp
 195 200 205
 5 Phe Lys Ala Ser Arg Val Tyr Thr Leu Val Leu Gly Phe Val Leu Pro
 210 215 220
 Val Cys Thr Ile Cys Val Leu Tyr Thr Asp Leu Leu Arg Arg Leu Arg
 225 230 235 240
 10 Ala Val Arg Leu Arg Ser Gly Ala Lys Ala Leu Gly Lys Ala Arg Arg
 245 250 255
 Lys Val Thr Val Leu Val Leu Val Val Leu Ala Val Cys Leu Leu Cys
 260 265 270
 Trp Thr Pro Phe His Leu Ala Ser Val Val Ala Leu Thr Thr Asp Leu
 275 280 285
 15 Pro Gln Thr Pro Leu Val Ile Ser Met Ser Tyr Val Ile Thr Ser Leu
 290 295 300
 Thr Tyr Ala Asn Ser Cys Leu Asn Pro Phe Leu Tyr Ala Phe Leu Asp
 305 310 315 320
 20 Asp Asn Phe Arg Lys Asn Phe Arg Ser Ile Leu Arg Cys
 325 330

(18) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 48 base pairs
 (B) TYPE: nucleic acid
 25 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

 (ii) MOLECULE TYPE: DNA (genomic)

 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

ACGAATTCAG CCATGGTCCT TGAGGTGAGT GACCACCAAG TGCTAAAT

48

30 (19) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 27 base pairs
 (B) TYPE: nucleic acid
 35 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

 (ii) MOLECULE TYPE: DNA (genomic)

 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GAGGATCCTG GAATGCGGGG AAGTCAG

27

(20) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

5

(A) LENGTH: 1107 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

10 ATGGTCCTTG AGGTGAGTGA CCACCAAGTG CTAAATGACG CCGAGGTTGC CGCCCTCCTG 60
GAGAACTTCA GCTCTTCCTA TGACTATGGA GAAAACGAGA GTGACTCGTG CTGTACCTCC 120
CCGCCCTGCC CACAGGACTT CAGCCTGAAC TTCGACCGGG CCTTCCTGCC AGCCCTCTAC 180
AGCCTCCTCT TTCTGCTGGG GCTGCTGGGC AACGGCGCGG TGGCAGCCGT GCTGCTGAGC 240
CGGCGGACAG CCCTGAGCAG CACCGACACC TTCCTGCTCC ACCTAGCTGT AGCAGACACG 300
15 CTGCTGGTGC TGACACTGCC GCTCTGGGCA GTGGACGCTG CCGTCCAGTG GGTCTTTGGC 360
TCTGGCCTCT GCAAAGTGGC AGGTGCCCTC TTCAACATCA ACTTCTACGC AGGAGCCCTC 420
CTGCTGGCCT GCATCAGCTT TGACCGCTAC CTGAACATAG TTCATGCCAC CCAGCTCTAC 480
CGCCGGGGGC CCCC GGCCCG CGTGACCCTC ACCTGCCTGG CTGTCTGGGG GCTCTGCCTG 540
CTTTTCGCCC TCCCAGACTT CATCTTCCTG TCGGCCACC ACGACGAGCG CCTCAACGCC 600
20 ACCCACTGCC AATACAACCT CCCACAGGTG GGCCGCACGG CTCTGCGGGT GCTGCAGCTG 660
GTGGCTGGCT TTCTGCTGCC CCTGCTGGTC ATGGCCTACT GCTATGCCCA CATCCTGGCC 720
GTGCTGCTGG TTTCCAGGGG CCAGCGGCGC CTGCGGGCCA TGCGGCTGGT GGTGGTGGTC 780
GTGGTGGCCT TTGCCCTCTG CTGGACCCCC TATCACCTGG TGGTGCTGGT GGACATCCTC 840
ATGGACCTGG GCGCTTTGGC CCGCAACTGT GGCCGAGAAA GCAGGGTAGA CGTGGCCAAG 900
25 TCGGTCACCT CAGGCCTGGG CTACATGCAC TGCTGCCTCA ACCCGCTGCT CTATGCCTTT 960
GTAGGGGTCA AGTTCCGGGA GCGGATGTGG ATGCTGCTCT TGCGCCTGGG CTGCCCCAAC 1020
CAGAGAGGGC TCCAGAGGCA GCCATCGTCT TCCGCGCGG ATTCACTCTG GTCTGAGACC 1080
TCAGAGGCCT CTTACTCGGG CTTGTGA 1107

(21) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 368 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

5

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Met Val Leu Glu Val Ser Asp His Gln Val Leu Asn Asp Ala Glu Val
 1 5 10 15

Ala Ala Leu Leu Glu Asn Phe Ser Ser Ser Tyr Asp Tyr Gly Glu Asn
 20 25 30

Glu Ser Asp Ser Cys Cys Thr Ser Pro Pro Cys Pro Gln Asp Phe Ser
 35 40 45

Leu Asn Phe Asp Arg Ala Phe Leu Pro Ala Leu Tyr Ser Leu Leu Phe
 50 55 60

Leu Leu Gly Leu Leu Gly Asn Gly Ala Val Ala Ala Val Leu Leu Ser
 65 70 75 80

Arg Arg Thr Ala Leu Ser Ser Thr Asp Thr Phe Leu Leu His Leu Ala
 85 90 95

Val Ala Asp Thr Leu Leu Val Leu Thr Leu Pro Leu Trp Ala Val Asp
 100 105 110

Ala Ala Val Gln Trp Val Phe Gly Ser Gly Leu Cys Lys Val Ala Gly
 115 120 125

Ala Leu Phe Asn Ile Asn Phe Tyr Ala Gly Ala Leu Leu Leu Ala Cys
 130 135 140

Ile Ser Phe Asp Arg Tyr Leu Asn Ile Val His Ala Thr Gln Leu Tyr
 145 150 155 160

Arg Arg Gly Pro Pro Ala Arg Val Thr Leu Thr Cys Leu Ala Val Trp
 165 170 175

Gly Leu Cys Leu Leu Phe Ala Leu Pro Asp Phe Ile Phe Leu Ser Ala
 180 185 190

His His Asp Glu Arg Leu Asn Ala Thr His Cys Gln Tyr Asn Phe Pro
 195 200 205

Gln Val Gly Arg Thr Ala Leu Arg Val Leu Gln Leu Val Ala Gly Phe
 210 215 220

Leu Leu Pro Leu Leu Val Met Ala Tyr Cys Tyr Ala His Ile Leu Ala
 225 230 235 240

	Val	Leu	Leu	Val	Ser	Arg	Gly	Gln	Arg	Arg	Leu	Arg	Ala	Met	Arg	Leu
					245					250					255	
	Val	Val	Val	Val	Val	Val	Ala	Phe	Ala	Leu	Cys	Trp	Thr	Pro	Tyr	His
				260					265					270		
5	Leu	Val	Val	Leu	Val	Asp	Ile	Leu	Met	Asp	Leu	Gly	Ala	Leu	Ala	Arg
			275					280					285			
	Asn	Cys	Gly	Arg	Glu	Ser	Arg	Val	Asp	Val	Ala	Lys	Ser	Val	Thr	Ser
		290					295					300				
10	Gly	Leu	Gly	Tyr	Met	His	Cys	Cys	Leu	Asn	Pro	Leu	Leu	Tyr	Ala	Phe
	305					310					315					320
	Val	Gly	Val	Lys	Phe	Arg	Glu	Arg	Met	Trp	Met	Leu	Leu	Leu	Arg	Leu
					325					330					335	
	Gly	Cys	Pro	Asn	Gln	Arg	Gly	Leu	Gln	Arg	Gln	Pro	Ser	Ser	Ser	Arg
				340					345					350		
15	Arg	Asp	Ser	Ser	Trp	Ser	Glu	Thr	Ser	Glu	Ala	Ser	Tyr	Ser	Gly	Leu
			355					360					365			

(22) INFORMATION FOR SEQ ID NO:21:

20

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

25 TTAAGCTTGA CCTAATGCCA TCTTGTGTCC 30

(23) INFORMATION FOR SEQ ID NO:22:

30

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

TTGGATCCAA AAGAACCATG CACCTCAGAG .30

35 (24) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1074 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

```

ATGGCTGATG ACTATGGCTC TGAATCCACA TCTTCCATGG AAGACTACGT TAACTTCAAC    60
TTCCTGACT TCTACTGTGA GAAAAACAAT GTCAGGCAGT TTGCGAGCCA TTCCTCCCA    120
CCCTTGTA CT GGCTCGTGT CATCGTGGGT GCCTTGGGCA ACAGTCTTGT TATCCTTGTC    180
10 TACTGGTACT GCACAAGAGT GAAGACCATG ACCGACATGT TCCTTTTGAA TTTGGCAATT    240
GCTGACCTCC TCTTTCTTGT CACTCTTCCC TTCTGGGCCA TTGCTGCTGC TGACCAGTGG    300
AAGTTCCAGA CCTTCATGTG CAAGGTGGTC AACAGCATGT ACAAGATGAA CTTCTACAGC    360
TGTGTGTTGC TGATCATGTG CATCAGCGTG GACAGGTACA TTGCCATTGC CCAGGCCATG    420
AGAGCACATA CTTGGAGGGA GAAAAGGCTT TTGTACAGCA AAATGGTTTG CTTTACCATC    480
15 TGGGTATTGG CAGCTGCTCT CTGCATCCCA GAAATCTTAT ACAGCCAAAT CAAGGAGGAA    540
TCCGGCATTG CTATCTGCAC CATGGTTTAC CCTAGCGATG AGAGCACCAA ACTGAAGTCA    600
GCTGTCTTGA CCCTGAAGGT CATTCTGGGG TTCTTCCTTC CCTTCGTGGT CATGGCTTGC    660
TGCTATACCA TCATCATTCA CACCCTGATA CAAGCCAAGA AGTCTTCCAA GCACAAAGCC    720
CTAAAAGTGA CCATCACTGT CCTGACCGTC TTTGTCTTGT CTCAGTTTCC CTACAACTGC    780
20 ATTTTGTGG TGCAGACCAT TGACGCCTAT GCCATGTTCA TCTCCAAC TGCCGTTTCC    840
ACCAACATTG ACATCTGCTT CCAGGTCACC CAGACCATCG CCTTCTTCCA CAGTTGCCTG    900
AACCCTGTTC TCTATGTTTT TGTGGGTGAG AGATTCCGCC GGGATCTCGT GAAAACCCTG    960
AAGAACTTGG GTTGCATCAG CCAGGCCAG TGGGTTTCAT TTACAAGGAG AGAGGGAAGC   1020
TTGAAGCTGT CGTCTATGTT GCTGGAGACA ACCTCAGGAG CACTCTCCCT CTGA       1074

```

25 (25) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 357 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant

30

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

	Met	Ala	Asp	Asp	Tyr	Gly	Ser	Glu	Ser	Thr	Ser	Ser	Met	Glu	Asp	Tyr	
	1				5					10					15		
5	Val	Asn	Phe	Asn	Phe	Thr	Asp	Phe	Tyr	Cys	Glu	Lys	Asn	Asn	Val	Arg	
				20					25					30			
	Gln	Phe	Ala	Ser	His	Phe	Leu	Pro	Pro	Leu	Tyr	Trp	Leu	Val	Phe	Ile	
			35					40					45				
	Val	Gly	Ala	Leu	Gly	Asn	Ser	Leu	Val	Ile	Leu	Val	Tyr	Trp	Tyr	Cys	
		50					55					60					
10	Thr	Arg	Val	Lys	Thr	Met	Thr	Asp	Met	Phe	Leu	Leu	Asn	Leu	Ala	Ile	
	65					70					75					80	
	Ala	Asp	Leu	Leu	Phe	Leu	Val	Thr	Leu	Pro	Phe	Trp	Ala	Ile	Ala	Ala	
					85					90					95		
15	Ala	Asp	Gln	Trp	Lys	Phe	Gln	Thr	Phe	Met	Cys	Lys	Val	Val	Asn	Ser	
				100						105					110		
	Met	Tyr	Lys	Met	Asn	Phe	Tyr	Ser	Cys	Val	Leu	Leu	Ile	Met	Cys	Ile	
			115					120					125				
	Ser	Val	Asp	Arg	Tyr	Ile	Ala	Ile	Ala	Gln	Ala	Met	Arg	Ala	His	Thr	
		130					135					140					
20	Trp	Arg	Glu	Lys	Arg	Leu	Leu	Tyr	Ser	Lys	Met	Val	Cys	Phe	Thr	Ile	
	145					150					155					160	
	Trp	Val	Leu	Ala	Ala	Ala	Leu	Cys	Ile	Pro	Glu	Ile	Leu	Tyr	Ser	Gln	
				165						170					175		
25	Ile	Lys	Glu	Glu	Ser	Gly	Ile	Ala	Ile	Cys	Thr	Met	Val	Tyr	Pro	Ser	
				180						185					190		
	Asp	Glu	Ser	Thr	Lys	Leu	Lys	Ser	Ala	Val	Leu	Thr	Leu	Lys	Val	Ile	
			195					200					205				
	Leu	Gly	Phe	Phe	Leu	Pro	Phe	Val	Val	Met	Ala	Cys	Cys	Tyr	Thr	Ile	
		210					215						220				
30	Ile	Ile	His	Thr	Leu	Ile	Gln	Ala	Lys	Lys	Ser	Ser	Lys	His	Lys	Ala	
	225					230					235					240	
	Leu	Lys	Val	Thr	Ile	Thr	Val	Leu	Thr	Val	Phe	Val	Leu	Ser	Gln	Phe	
				245						250					255		
35	Pro	Tyr	Asn	Cys	Ile	Leu	Leu	Val	Gln	Thr	Ile	Asp	Ala	Tyr	Ala	Met	
				260					265					270			
	Phe	Ile	Ser	Asn	Cys	Ala	Val	Ser	Thr	Asn	Ile	Asp	Ile	Cys	Phe	Gln	
			275					280					285				

Val Thr Gln Thr Ile Ala Phe Phe His Ser Cys Leu Asn Pro Val Leu
 290 295 300

Tyr Val Phe Val Gly Glu Arg Phe Arg Arg Asp Leu Val Lys Thr Leu
 305 310 315 320

5 Lys Asn Leu Gly Cys Ile Ser Gln Ala Gln Trp Val Ser Phe Thr Arg
 325 330 335

Arg Glu Gly Ser Leu Lys Leu Ser Ser Met Leu Leu Glu Thr Thr Ser
 340 345 350

Gly Ala Leu Ser Leu
 10 355

(26) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1110 base pairs
 (B) TYPE: nucleic acid
 15 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

ATGGCCTCAT CGACCACTCG GGGCCCCAGG GTTCTGACT TATTTTCTGG GCTGCCGCCG 60

20 GCGGTCACAA CTCCGCCCAA CCAGAGCGCA GAGGCCTCGG CGGGCAACGG GTCGGTGGCT 120

GGCGCGGACG CTCCAGCCGT CACGCCCTTC CAGAGCCTGC AGCTGGTGCA TCAGCTGAAG 180

GGGCTGATCG TGCTGCTCTA CAGCGTCGTG GTGGTCGTGG GGCTGGTGGG CAACTGCCTG 240

CTGGTGCTGG TGATCGCGCG GGTGCCGCGG CTGCACAACG TGACGAACTT CCTCATCGGC 300

AACCTGGCCT TGTCCGACGT GTCATGTGC ACCGCCTGCG TGCCGCTCAC GCTGGCCTAT 360

25 GCCTTCGAGC CACGCGGCTG GGTGTTCCGG GCGGCCTGT GCCACCTGGT CTTCTTCCTG 420

CAGCCGGTCA CCGTCTATGT GTCGGTGTTC ACGCTACCA CCATCGCAGT GGACCGCTAC 480

GTCGTGCTGG TGCACCCGCT GAGGCGCGCA TCTCGCTGCG CCTCAGCCTA CGCTGTGCTG 540

GCCATCTGGG CGCTGTCCGC GGTGCTGGCG CTGCCGCCCG CCGTGACAC CTATCACGTG 600

GAGCTCAAGC CGCAGACGT GCGCCTCTGC GAGGAGTTCT GGGGCTCCCA GGAGCGCCAG 660

30 CGCCAGCTCT ACGCCTGGGG GCTGCTGCTG GTCACCTACC TGCTCCCTCT GCTGGTCATC 720

CTCCTGTCTT ACGTCCGGGT GTCAGTGAAG CTCCGCAACC GCGTGGTGCC GGGCTGCGTG 780

ACCCAGAGCC AGGCCGACTG GGACCGCGCT CGGCGCCGGC GCACCTTCTG CTTGCTGGTG 840

GTGGTCGTGG TGGTGTTCGC CGTCTGCTGG CTGCCGCTGC ACGTCTTCAA CCTGCTGCGG 900
 GACCTCGACC CCCACGCCAT CGACCCTTAC GCCTTTGGGC TGGTGCAGCT GCTCTGCCAC 960
 TGGCTCGCCA TGAGTTCGGC CTGCTACAAC CCCTTCATCT ACCCCTGGCT GCACGACAGC 1020
 TTCCGCGAGG AGCTGCGCAA ACTGTTGGTC GCTTGGCCCC GCAAGATAGC CCCCCATGGC 1080
 5 CAGAATATGA CCGTCAGCGT GGTCATCTGA 1110

(27) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 369 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

15 Met Ala Ser Ser Thr Thr Arg Gly Pro Arg Val Ser Asp Leu Phe Ser
 1 5 10 15
 Gly Leu Pro Pro Ala Val Thr Thr Pro Ala Asn Gln Ser Ala Glu Ala
 20 20 25 30
 Ser Ala Gly Asn Gly Ser Val Ala Gly Ala Asp Ala Pro Ala Val Thr
 35 40 45
 20 Pro Phe Gln Ser Leu Gln Leu Val His Gln Leu Lys Gly Leu Ile Val
 50 55 60
 Leu Leu Tyr Ser Val Val Val Val Val Gly Leu Val Gly Asn Cys Leu
 65 70 75 80
 25 Leu Val Leu Val Ile Ala Arg Val Pro Arg Leu His Asn Val Thr Asn
 85 90 95
 Phe Leu Ile Gly Asn Leu Ala Leu Ser Asp Val Leu Met Cys Thr Ala
 100 105 110
 Cys Val Pro Leu Thr Leu Ala Tyr Ala Phe Glu Pro Arg Gly Trp Val
 115 120 125
 30 Phe Gly Gly Gly Leu Cys His Leu Val Phe Phe Leu Gln Pro Val Thr
 130 135 140
 Val Tyr Val Ser Val Phe Thr Leu Thr Thr Ile Ala Val Asp Arg Tyr
 145 150 155 160
 35 Val Val Leu Val His Pro Leu Arg Arg Ala Ser Arg Cys Ala Ser Ala
 165 170 175

Tyr Ala Val Leu Ala Ile Trp Ala Leu Ser Ala Val Leu Ala Leu Pro
 180 185 190
 Pro Ala Val His Thr Tyr His Val Glu Leu Lys Pro His Asp Val Arg
 195 200 205
 5 Leu Cys Glu Glu Phe Trp Gly Ser Gln Glu Arg Gln Arg Gln Leu Tyr
 210 215 220
 Ala Trp Gly Leu Leu Leu Val Thr Tyr Leu Leu Pro Leu Leu Val Ile
 225 230 235 240
 10 Leu Leu Ser Tyr Val Arg Val Ser Val Lys Leu Arg Asn Arg Val Val
 245 250 255
 Pro Gly Cys Val Thr Gln Ser Gln Ala Asp Trp Asp Arg Ala Arg Arg
 260 265 270
 Arg Arg Thr Phe Cys Leu Leu Val Val Val Val Val Phe Ala Val
 275 280 285
 15 Cys Trp Leu Pro Leu His Val Phe Asn Leu Leu Arg Asp Leu Asp Pro
 290 295 300
 His Ala Ile Asp Pro Tyr Ala Phe Gly Leu Val Gln Leu Leu Cys His
 305 310 315 320
 20 Trp Leu Ala Met Ser Ser Ala Cys Tyr Asn Pro Phe Ile Tyr Ala Trp
 325 330 335
 Leu His Asp Ser Phe Arg Glu Glu Leu Arg Lys Leu Leu Val Ala Trp
 340 345 350
 Pro Arg Lys Ile Ala Pro His Gly Gln Asn Met Thr Val Ser Val Val
 355 360 365
 25 Ile

(28) INFORMATION FOR SEQ ID NO:27:

30 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1083 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

35 ATGGACCCAG AAGAACTTC AGTTTATTG GATTATTACT ATGCTACGAG CCCAACTCT 60
 GACATCAGGG AGACCCACTC CCATGTTCTT TACACCTCTG TCTTCCTTCC AGTCTTTTAC 120

ACAGCTGTGT TCCTGACTGG AGTGCTGGGG AACCTTGTTT TCATGGGAGC GTTGCATTTT 180
 AAACCCGGCA GCCGAAGACT GATCGACATC TTTATCATCA ATCTGGCTGC CTCTGACTTC 240
 ATTTTCTTGG TCACATTGCC TCTCTGGGTG GATAAAGAAG CATCTCTAGG ACTGTGGAGG 300
 ACGGGCTCCT TCCTGTGCAA AGGGAGCTCC TACATGATCT CCGTCAATAT GCACTGCAGT 360
 5 GTCCTCCTGC TCACTTGCAT GAGTGTGAC CGCTACCTGG CCATTGTGTG GCCAGTCGTA 420
 TCCAGGAAAT TCAGAAGGAC AGACTGTGCA TATGTAGTCT GTGCCAGCAT CTGGTTTATC 480
 TCCTGCCTGC TGGGGTTGCC TACTCTTCTG TCCAGGGAGC TCACGCTGAT TGATGATAAG 540
 CCATACTGTG CAGAGAAAAA GGCAACTCCA ATTAAACTCA TATGGTCCCT GGTGGCCTTA 600
 ATTTTCACCT TTTTGTCCC TTTGTTGAGC ATTGTGACCT GCTACTGTTG CATTGCAAGG 660
 10 AAGCTGTGTG CCCATTACCA GCAATCAGGA AAGCACAACA AAAAGCTGAA GAAATCTATA 720
 AAGATCATCT TTATTGTCGT GGCAGCCTTT CTTGTCTCCT GGCTGCCCTT CAATACTTTC 780
 AAGTTCCTGG CCATTGTCTC TGGGTTGCGG CAAGAACACT ATTTACCCTC AGCTATTCTT 840
 CAGCTTG GTA TGGAGGTGAG TGGACCCTTG GCATTTGCCA ACAGCTGTGT CAACCCTTTC 900
 ATTTACTATA TCTTCGACAG CTACATCCGC CGGGCCATTG TCCACTGCTT GTGCCCTTGC 960
 15 CTGAAAAACT ATGACTTTGG GAGTAGCACT GAGACATCAG ATAGTCACCT CACTAAGGCT 1020
 CTCTCCACCT TCATTCATGC AGAAGATTTT GCCAGGAGGA GGAAGAGGTC TGTGTCACTC 1080
 TAA 1083

(29) INFORMATION FOR SEQ ID NO:28:

- 20 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 360 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

- 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

Met Asp Pro Glu Glu Thr Ser Val Tyr Leu Asp Tyr Tyr Tyr Ala Thr
 1 5 10 15
 Ser Pro Asn Ser Asp Ile Arg Glu Thr His Ser His Val Pro Tyr Thr
 20 25 30
 30 Ser Val Phe Leu Pro Val Phe Tyr Thr Ala Val Phe Leu Thr Gly Val
 35 40 45

	Leu Gly Asn Leu Val Leu Met Gly Ala Leu His Phe Lys Pro Gly Ser	50	55	60
	Arg Arg Leu Ile Asp Ile Phe Ile Ile Asn Leu Ala Ala Ser Asp Phe	65	70	75 80
5	Ile Phe Leu Val Thr Leu Pro Leu Trp Val Asp Lys Glu Ala Ser Leu	85	90	95
	Gly Leu Trp Arg Thr Gly Ser Phe Leu Cys Lys Gly Ser Ser Tyr Met	100	105	110
10	Ile Ser Val Asn Met His Cys Ser Val Leu Leu Leu Thr Cys Met Ser	115	120	125
	Val Asp Arg Tyr Leu Ala Ile Val Trp Pro Val Val Ser Arg Lys Phe	130	135	140
	Arg Arg Thr Asp Cys Ala Tyr Val Val Cys Ala Ser Ile Trp Phe Ile	145	150	155 160
15	Ser Cys Leu Leu Gly Leu Pro Thr Leu Leu Ser Arg Glu Leu Thr Leu	165	170	175
	Ile Asp Asp Lys Pro Tyr Cys Ala Glu Lys Lys Ala Thr Pro Ile Lys	180	185	190
20	Leu Ile Trp Ser Leu Val Ala Leu Ile Phe Thr Phe Phe Val Pro Leu	195	200	205
	Leu Ser Ile Val Thr Cys Tyr Cys Cys Ile Ala Arg Lys Leu Cys Ala	210	215	220
	His Tyr Gln Gln Ser Gly Lys His Asn Lys Lys Leu Lys Lys Ser Ile	225	230	235 240
25	Lys Ile Ile Phe Ile Val Val Ala Ala Phe Leu Val Ser Trp Leu Pro	245	250	255
	Phe Asn Thr Phe Lys Phe Leu Ala Ile Val Ser Gly Leu Arg Gln Glu	260	265	270
30	His Tyr Leu Pro Ser Ala Ile Leu Gln Leu Gly Met Glu Val Ser Gly	275	280	285
	Pro Leu Ala Phe Ala Asn Ser Cys Val Asn Pro Phe Ile Tyr Tyr Ile	290	295	300
	Phe Asp Ser Tyr Ile Arg Arg Ala Ile Val His Cys Leu Cys Pro Cys	305	310	315 320
35	Leu Lys Asn Tyr Asp Phe Gly Ser Ser Thr Glu Thr Ser Asp Ser His	325	330	335
	Leu Thr Lys Ala Leu Ser Thr Phe Ile His Ala Glu Asp Phe Ala Arg			

340

345

350

Arg Arg Lys Arg Ser Val Ser Leu
355 360

(30) INFORMATION FOR SEQ ID NO:29:

- 5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

- 10 (ii) MOLECULE TYPE: DNA (genomic)

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

CTAGAATTCT GACTCCAGCC AAAGCATGAA T

31

(31) INFORMATION FOR SEQ ID NO:30:

- 15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA (genomic)

- 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GCTGGATCCT AAACAGTCTG CGCTCGGCCT

30

(32) INFORMATION FOR SEQ ID NO:31:

- 25 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1020 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA (genomic)

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

30 ATGAATGGCC TTGAAGTGGC TCCCCCAGGT CTGATCACCA ACTTCTCCCT GGCCACGGCA 60
GAGCAATGTG GCCAGGAGAC GCCACTGGAG AACATGCTGT TCGCCTCCTT CTACCTTCTG 120
GATTTTATCC TGGCTTTAGT TGGCAATACC CTGGCTCTGT GGCTTTTCAT CCGAGACCAC 180
AAGTCCGGGA CCCC GGCCAA CGTGTTCTCTG ATGCATCTGG CCGTGGCCGA CTTGTCGTGC 240
GTGCTGGTCC TGCCCCACCCG CCTGGTCTAC CACTTCTCTG GGAACCACTG GCCATTTGGG 300

GAAATCGCAT GCCGTCTCAC CGGCTTCTC TTCTACCTCA ACATGTACGC CAGCATCTAC 360
 TTCTCACCT GCATCAGCGC CGACCGTTTC CTGGCCATTG TGCACCCGGT CAAGTCCCTC 420
 AAGCTCCGCA GGCCCTCTA CGCACACCTG GCCTGTGCCT TCCTGTGGGT GGTGGTGGCT 480
 GTGGCCATGG CCCCCTGCT GGTGAGCCCA CAGACCGTGC AGACCAACCA CACGGTGGTC 540
 5 TGCCTGCAGC TGTACCGGGA GAAGGCCTCC CACCATGCCC TGGTGTCCCT GGCAGTGGCC 600
 TTCACCTTCC CGTTCATCAC CACGGTCACC TGCTACCTGC TGATCATCCG CAGCCTGCGG 660
 CAGGGCCTGC GTGTGGAGAA GCGCCTCAAG ACCAAGGCAG TGCGCATGAT CGCCATAGTG 720
 CTGGCCATCT TCCTGGTCTG CTTCGTGCCC TACCACGTCA ACCGCTCCGT CTACGTGCTG 780
 CACTACCGCA GCCATGGGGC CTCCTGCGCC ACCCAGCGCA TCCTGGCCCT GGCAAACCGC 840
 10 ATCACCTCCT GCCTCACCAG CCTCAACGGG GCACTCGACC CCATCATGTA TTTCTTCGTG 900
 GCTGAGAAGT TCCGCCACGC CCTGTGCAAC TTGCTCTGTG GCAAAGGCT CAAGGGCCCG 960
 CCCCCAGCT TCGAAGGGAA AACCAACGAG AGCTCGCTGA GTGCCAAGTC AGAGCTGTGA 1020

(33) INFORMATION FOR SEQ ID NO:32:

15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 339 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

Met Asn Gly Leu Glu Val Ala Pro Pro Gly Leu Ile Thr Asn Phe Ser
 1 5 10 15

Leu Ala Thr Ala Glu Gln Cys Gly Gln Glu Thr Pro Leu Glu Asn Met
 20 25 30

25 Leu Phe Ala Ser Phe Tyr Leu Leu Asp Phe Ile Leu Ala Leu Val Gly
 35 40 45

Asn Thr Leu Ala Leu Trp Leu Phe Ile Arg Asp His Lys Ser Gly Thr
 50 55 60

30 Pro Ala Asn Val Phe Leu Met His Leu Ala Val Ala Asp Leu Ser Cys
 65 70 75 80

Val Leu Val Leu Pro Thr Arg Leu Val Tyr His Phe Ser Gly Asn His
 85 90 95

Trp Pro Phe Gly Glu Ile Ala Cys Arg Leu Thr Gly Phe Leu Phe Tyr

27

	100	105	110
	Leu Asn Met Tyr Ala Ser Ile Tyr Phe Leu Thr Cys Ile Ser Ala Asp 115 120 125		
5	Arg Phe Leu Ala Ile Val His Pro Val Lys Ser Leu Lys Leu Arg Arg 130 135 140		
	Pro Leu Tyr Ala His Leu Ala Cys Ala Phe Leu Trp Val Val Val Ala 145 150 155 160		
	Val Ala Met Ala Pro Leu Leu Val Ser Pro Gln Thr Val Gln Thr Asn 165 170 175		
10	His Thr Val Val Cys Leu Gln Leu Tyr Arg Glu Lys Ala Ser His His 180 185 190		
	Ala Leu Val Ser Leu Ala Val Ala Phe Thr Phe Pro Phe Ile Thr Thr 195 200 205		
15	Val Thr Cys Tyr Leu Leu Ile Ile Arg Ser Leu Arg Gln Gly Leu Arg 210 215 220		
	Val Glu Lys Arg Leu Lys Thr Lys Ala Val Arg Met Ile Ala Ile Val 225 230 235 240		
	Leu Ala Ile Phe Leu Val Cys Phe Val Pro Tyr His Val Asn Arg Ser 245 250 255		
20	Val Tyr Val Leu His Tyr Arg Ser His Gly Ala Ser Cys Ala Thr Gln 260 265 270		
	Arg Ile Leu Ala Leu Ala Asn Arg Ile Thr Ser Cys Leu Thr Ser Leu 275 280 285		
25	Asn Gly Ala Leu Asp Pro Ile Met Tyr Phe Phe Val Ala Glu Lys Phe 290 295 300		
	Arg His Ala Leu Cys Asn Leu Leu Cys Gly Lys Arg Leu Lys Gly Pro 305 310 315 320		
	Pro Pro Ser Phe Glu Gly Lys Thr Asn Glu Ser Ser Leu Ser Ala Lys 325 330 335		
30	Ser Glu Leu		

(34) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 29 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

ATAAGATGAT CACCCTGAAC AATCAAGAT

29

(35) INFORMATION FOR SEQ ID NO:34:

5

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 33 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

TCCGAATTCA TAACATTTCa CTGTTTATAT TGC

33

(36) INFORMATION FOR SEQ ID NO:35:

15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 996 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

20

ATGATCACCC TGAACAATCA AGATCAACCT GTCACCTTTTA ACAGCTCACA TCCAGATGAA 60

TACAAAATTG CAGCCCTTGT CTTCTATAGC TGTATCTTCA TAATTGGATT ATTTGTTAAC 120

ATCACTGCAT TATGGGTTTT CAGTTGTACC ACCAAGAAGA GAACCACGGT AACCATCTAT 180

ATGATGAATG TGGCATTAGT GGACTTGATA TTTATAATGA CTTTACCCTT TCGAATGTTT 240

TATTATGCAA AAGATGCATG GCCATTTGGA GAGTACTTCT GCCAGATTAT TGGAGCTCTC 300

25

ACAGTGTTTT ACCCAAGCAT TGCTTTATGG CTTCTTGCCT TTATTAGTGC TGACAGATAC 360

ATGGCCATTG TACAGCCGAA GTACGCCAAA GAACTTAAAA ACACGTGCAA AGCCGTGCTG 420

GCGTGTGTGG GAGTCTGGAT AATGACCCTG ACCACGACCA CCCCTCTGCT ACTGCTCTAT 480

AAAGACCCAG ATAAAGACTC CACTCCCGCC ACCTGCCTCA AGATTTCTGA CATCATCTAT 540

CTAAAAGCTG TGAACGTGCT GAACCTCACT CGACTGACAT TTTTTTCTT GATTCCTTTG 600

30

TTCATCATGA TTGGGTGCTA CTTGGTCATT ATTCATAATC TCCTTCACGG CAGGACGTCT 660

AAGCTGAAAC CCAAAGTCAA GGAGAAGTCC ATAAGGATCA TCATCACGCT GCTGGTGCAG 720

GTGCTCGTCT GCTTTATGCC CTTCCACATC TGTTTCGCTT TCCTGATGCT GGGAACGGGG 780
 GAGAACAGTT ACAATCCCTG GGGAGCCTTT ACCACCTTCC TCATGAACCT CAGCACGTGT 840
 CTGGATGTGA TTCTCTACTA CATCGTTTCA AAACAATTTT AGGCTCGAGT CATTAGTGTC 900
 ATGCTATAACC GTAATTACCT TCGAAGCCTG CGCAGAAAAA GTTTCCGATC TGGTAGTCTA 960
 5 AGGTCACTAA GCAATATAAA CAGTGAAATG TTATGA 996

(37) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 331 amino acids

(B) TYPE: amino acid

10 (C) STRANDEDNESS:

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

15 Met Ile Thr Leu Asn Asn Gln Asp Gln Pro Val Thr Phe Asn Ser Ser
 1 5 10 15
 His Pro Asp Glu Tyr Lys Ile Ala Ala Leu Val Phe Tyr Ser Cys Ile
 20 20 25 30
 Phe Ile Ile Gly Leu Phe Val Asn Ile Thr Ala Leu Trp Val Phe Ser
 35 40 45
 20 Cys Thr Thr Lys Lys Arg Thr Thr Val Thr Ile Tyr Met Met Asn Val
 50 55 60
 Ala Leu Val Asp Leu Ile Phe Ile Met Thr Leu Pro Phe Arg Met Phe
 65 70 75 80
 25 Tyr Tyr Ala Lys Asp Ala Trp Pro Phe Gly Glu Tyr Phe Cys Gln Ile
 85 90 95
 Ile Gly Ala Leu Thr Val Phe Tyr Pro Ser Ile Ala Leu Trp Leu Leu
 100 105 110
 Ala Phe Ile Ser Ala Asp Arg Tyr Met Ala Ile Val Gln Pro Lys Tyr
 115 120 125
 30 Ala Lys Glu Leu Lys Asn Thr Cys Lys Ala Val Leu Ala Cys Val Gly
 130 135 140
 Val Trp Ile Met Thr Leu Thr Thr Thr Thr Pro Leu Leu Leu Tyr
 145 150 155 160
 35 Lys Asp Pro Asp Lys Asp Ser Thr Pro Ala Thr Cys Leu Lys Ile Ser
 165 170 175

	Asp	Ile	Ile	Tyr	Leu	Lys	Ala	Val	Asn	Val	Leu	Asn	Leu	Thr	Arg	Leu	
					180				185					190			
	Thr	Phe	Phe	Phe	Leu	Ile	Pro	Leu	Phe	Ile	Met	Ile	Gly	Cys	Tyr	Leu	
					195			200					205				
5	Val	Ile	Ile	His	Asn	Leu	Leu	His	Gly	Arg	Thr	Ser	Lys	Leu	Lys	Pro	
		210				215					220						
	Lys	Val	Lys	Glu	Lys	Ser	Ile	Arg	Ile	Ile	Ile	Thr	Leu	Leu	Val	Gln	
	225				230					235					240		
10	Val	Leu	Val	Cys	Phe	Met	Pro	Phe	His	Ile	Cys	Phe	Ala	Phe	Leu	Met	
				245					250						255		
	Leu	Gly	Thr	Gly	Glu	Asn	Ser	Tyr	Asn	Pro	Trp	Gly	Ala	Phe	Thr	Thr	
				260				265						270			
	Phe	Leu	Met	Asn	Leu	Ser	Thr	Cys	Leu	Asp	Val	Ile	Leu	Tyr	Tyr	Ile	
			275				280						285				
15	Val	Ser	Lys	Gln	Phe	Gln	Ala	Arg	Val	Ile	Ser	Val	Met	Leu	Tyr	Arg	
		290				295						300					
	Asn	Tyr	Leu	Arg	Ser	Leu	Arg	Arg	Lys	Ser	Phe	Arg	Ser	Gly	Ser	Leu	
	305				310					315					320		
20	Arg	Ser	Leu	Ser	Asn	Ile	Asn	Ser	Glu	Met	Leu						
				325				330									

(38) INFORMATION FOR SEQ ID NO:37:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 28 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

CCAAGCTTCC AGGCCTGGGG TGTGCTGG

28

30 (39) INFORMATION FOR SEQ ID NO:38:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 29 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

35

ATGGATCCTG ACCTTCGGCC CCTGGCAGA

29

(40) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1077 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

10 ATGCCCTCTG TGTCTCCAGC GGGGCCCTCG GCCGGGGCAG TCCCCAATGC CACCGCAGTG 60
 ACAACAGTGC GGACCAATGC CAGCGGGCTG GAGGTGCCCC GTTTCACCT GTTTGCCCCG 120
 CTGGACGAGG AGCTGCATGG CACCTTCCCA GGCCTGTGCG TGGCGCTGAT GGCGGTGCAC 180
 GGAGCCATCT TCCTGGCAGG GCTGGTGCTC AACGGGCTGG CGCTGTACGT CTTCTGCTGC 240
 CGCACCCGGG CCAAGACACC CTCAGTCATC TACACCATCA ACCTGGTGGT GACCGATCTA 300
 15 CTGGTAGGGC TGTCCCTGCC CACGCGCTTC GCTGTGTACT ACGGCGCCAG GGGCTGCCTG 360
 CGCTGTGCCT TCCCGCACGT CCTCGGTTAC TTCCTCAACA TGCACTGCTC CATCCTCTTC 420
 CTCACCTGCA TCTGCGTGGA CCGCTACCTG GCCATCGTGC GGCCCGAAGG CTCCCGCCGC 480
 TGCCGCCAGC CTGCCTGTGC CAGGGCCGTG TGCGCCTTCG TGTGGCTGGC CGCCGGTGCC 540
 GTCACCCTGT CGGTGCTGGG CGTGACAGGC AGCCGGCCCT GCTGCCGTGT CTTTGCCTG 600
 20 ACTGTCCTGG AGTTCCTGCT GCCCCTGCTG GTCATCAGCG TGTTTACCGG CCGCATCATG 660
 TGTGCACTGT CGCGGCCGGG TCTGCTCCAC CAGGGTCGCC AGCGCCGCGT GCGGGCCATG 720
 CAGCTCCTGC TCACGGTGCT CATCATCTTT CTCGTCTGCT TCACGCCCTT CCACGCCCCG 780
 CAAGTGGCCG TGGCGCTGTG GCCCGACATG CCACACCACA CGAGCCTCGT GGTCTACCAC 840
 GTGGCCGTGA CCCTCAGCAG CCTCAACAGC TGCATGGACC CCATCGTCTA CTGCTTCGTC 900
 25 ACCAGTGGCT TCCAGGCCAC CGTCCGAGGC CTCTTCGGCC AGCACGGAGA GCGTGAGCCC 960
 AGCAGCGGTG ACGTGGTCAG CATGCACAGG AGCTCCAAGG GCTCAGGCCG TCATCACATC 1020
 CTCAGTGCCG GCCCTCACGC CCTCACCAG GCCCTGGCTA ATGGGCCCCGA GGCTTAG 1077

(41) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 358 amino acids

(B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

	Met	Pro	Ser	Val	Ser	Pro	Ala	Gly	Pro	Ser	Ala	Gly	Ala	Val	Pro	Asn	
	1				5				10						15		
	Ala	Thr	Ala	Val	Thr	Thr	Val	Arg	Thr	Asn	Ala	Ser	Gly	Leu	Glu	Val	
				20				25						30			
10	Pro	Leu	Phe	His	Leu	Phe	Ala	Arg	Leu	Asp	Glu	Glu	Leu	His	Gly	Thr	
			35					40					45				
	Phe	Pro	Gly	Leu	Cys	Val	Ala	Leu	Met	Ala	Val	His	Gly	Ala	Ile	Phe	
		50					55					60					
15	Leu	Ala	Gly	Leu	Val	Leu	Asn	Gly	Leu	Ala	Leu	Tyr	Val	Phe	Cys	Cys	
	65				70					75					80		
	Arg	Thr	Arg	Ala	Lys	Thr	Pro	Ser	Val	Ile	Tyr	Thr	Ile	Asn	Leu	Val	
					85					90				95			
	Val	Thr	Asp	Leu	Leu	Val	Gly	Leu	Ser	Leu	Pro	Thr	Arg	Phe	Ala	Val	
				100				105						110			
20	Tyr	Tyr	Gly	Ala	Arg	Gly	Cys	Leu	Arg	Cys	Ala	Phe	Pro	His	Val	Leu	
			115					120					125				
	Gly	Tyr	Phe	Leu	Asn	Met	His	Cys	Ser	Ile	Leu	Phe	Leu	Thr	Cys	Ile	
		130					135					140					
25	Cys	Val	Asp	Arg	Tyr	Leu	Ala	Ile	Val	Arg	Pro	Glu	Ala	Pro	Ala	Ala	
	145					150				155					160		
	Cys	Arg	Gln	Pro	Ala	Cys	Ala	Arg	Ala	Val	Cys	Ala	Phe	Val	Trp	Leu	
				165						170					175		
	Ala	Ala	Gly	Ala	Val	Thr	Leu	Ser	Val	Leu	Gly	Val	Thr	Gly	Ser	Arg	
				180				185						190			
30	Pro	Cys	Cys	Arg	Val	Phe	Ala	Leu	Thr	Val	Leu	Glu	Phe	Leu	Leu	Pro	
			195					200					205				
	Leu	Leu	Val	Ile	Ser	Val	Phe	Thr	Gly	Arg	Ile	Met	Cys	Ala	Leu	Ser	
		210					215					220					
35	Arg	Pro	Gly	Leu	Leu	His	Gln	Gly	Arg	Gln	Arg	Arg	Val	Arg	Ala	Met	
	225					230				235					240		
	Gln	Leu	Leu	Leu	Thr	Val	Leu	Ile	Ile	Phe	Leu	Val	Cys	Phe	Thr	Pro	
				245						250					255		

Phe His Ala Arg Gln Val Ala Val Ala Leu Trp Pro Asp Met Pro His
260 265 270

His Thr Ser Leu Val Val Tyr His Val Ala Val Thr Leu Ser Ser Leu
275 280 285

[illegible]

Gln Ala Thr Val Arg Gly Leu Phe Gly Gln His Gly Glu Arg Glu Pro
305 310 315 320

10 Ser Ser Gly Asp Val Val Ser Met His Arg Ser Ser Lys Gly Ser Gly
 325 330 335

Arg His His Ile Leu Ser Ala Gly Pro His Ala Leu Thr Gln Ala Leu
340 345 350

Ala Asn Gly Pro Glu Ala
355

15 (42) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GAGAATTCAC TCCTGAGCTC AAGATGAACT

30

(43) INFORMATION FOR SEQ ID NO:42:

25 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

30 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

CGGGATCCCC GTAAGTGGAGC CACTTCAGAT

30

(44) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1050 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

```

ATGAACTCCA CCTTGGATGG TAATCAGAGC AGCCACCCTT TTTGCCTCTT GGCATTGTCG 60
5  TATTTGGAAA CTGTCAATTT TTGCCTTTTG GAAGTATTGA TTATTGTCTT TCTAACTGTA 120
    TTGATTATTT CTGGCAACAT CATTGTGATT TTTGTATTTT ACTGTGCACC TTTGTTGAAC 180
    CATCACACTA CAAGTTATTT TATCCAGACT ATGGCATATG CTGACCTTTT TGTGGGGGTG 240
    AGCTGCGTGG TCCCTTCTTT ATCACTCCTC CATCACCCCC TTCCAGTAGA GGAGTCCTTG 300
    ACTTGCCAGA TATTTGGTTT TGTAGTATCA GTTCTGAAGA GCGTCTCCAT GGCTTCTCTG 360
10 GCCTGTATCA GCATTGATAG ATACATTGCC ATTACTAAAC CTTTAACCTA TAATACTCTG 420
    GTTACACCCT GGAGACTACG CCTGTGTATT TTCCTGATTT GGCTATACTC GACCCTGGTC 480
    TTCCTGCCTT CCTTTTCCA CTGGGGCAA CCGTGATATC ATGGAGATGT GTTTCAGTGG 540
    TGTGCGGAGT CCTGGCACAC CGACTCCTAC TTCACCCTGT TCATCGTGAT GATGTTATAT 600
    GCCCCAGCAG CCCTTATTGT CTGCTTCACC TATTTCAACA TCTCCGCAT CTGCCAACAG 660
15 CACACAAAGG ATATCAGCGA AAGGCAAGCC CGCTTCAGCA GCCAGAGTGG GGAGACTGGG 720
    GAAGTGCAGG CCTGTCTGA TAAGCGCTAT GCCATGGTCC TGTTTCGAAT CACTAGTGTA 780
    TTTTACATCC TCTGGTTGCC ATATATCATC TACTTCTTGT TGGAAAGCTC CACTGGCCAC 840
    AGCAACCGCT TCGCATCCTT CTTGACCACC TGGCTTGCTA TTAGTAACAG TTTCTGCAAC 900
    TGTGTAATTT ATAGTCTCTC CAACAGTGTA TTCCAAGAG GACTAAAGCG CCTCTCAGGG 960
20 GCTATGTGTA CTTCTGTGC AAGTCAGACT ACAGCCAACG ACCCTTACAC AGTTAGAAGC 1020
    AAAGGCCCTC TTAATGGATG TCATATCTGA 1050

```

(45) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:

- 25 (A) LENGTH: 349 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

30 Met Asn Ser Thr Leu Asp Gly Asn Gln Ser Ser His Pro Phe Cys Leu

35

	1		5		10		15									
	Leu	Ala	Phe	Gly	Tyr	Leu	Glu	Thr	Val	Asn	Phe	Cys	Leu	Leu	Glu	Val
				20					25						30	
5	Leu	Ile	Ile	Val	Phe	Leu	Thr	Val	Leu	Ile	Ile	Ser	Gly	Asn	Ile	Ile
			35					40					45			
	Val	Ile	Phe	Val	Phe	His	Cys	Ala	Pro	Leu	Leu	Asn	His	His	Thr	Thr
		50					55					60				
	Ser	Tyr	Phe	Ile	Gln	Thr	Met	Ala	Tyr	Ala	Asp	Leu	Phe	Val	Gly	Val
	65					70					75				80	
10	Ser	Cys	Val	Val	Pro	Ser	Leu	Ser	Leu	Leu	His	His	Pro	Leu	Pro	Val
					85					90					95	
	Glu	Glu	Ser	Leu	Thr	Cys	Gln	Ile	Phe	Gly	Phe	Val	Val	Ser	Val	Leu
				100					105					110		
15	Lys	Ser	Val	Ser	Met	Ala	Ser	Leu	Ala	Cys	Ile	Ser	Ile	Asp	Arg	Tyr
			115					120					125			
	Ile	Ala	Ile	Thr	Lys	Pro	Leu	Thr	Tyr	Asn	Thr	Leu	Val	Thr	Pro	Trp
		130					135					140				
	Arg	Leu	Arg	Leu	Cys	Ile	Phe	Leu	Ile	Trp	Leu	Tyr	Ser	Thr	Leu	Val
	145				150					155					160	
20	Phe	Leu	Pro	Ser	Phe	Phe	His	Trp	Gly	Lys	Pro	Gly	Tyr	His	Gly	Asp
					165					170					175	
	Val	Phe	Gln	Trp	Cys	Ala	Glu	Ser	Trp	His	Thr	Asp	Ser	Tyr	Phe	Thr
				180					185					190		
25	Leu	Phe	Ile	Val	Met	Met	Leu	Tyr	Ala	Pro	Ala	Ala	Leu	Ile	Val	Cys
			195					200					205			
	Phe	Thr	Tyr	Phe	Asn	Ile	Phe	Arg	Ile	Cys	Gln	Gln	His	Thr	Lys	Asp
		210					215					220				
	Ile	Ser	Glu	Arg	Gln	Ala	Arg	Phe	Ser	Ser	Gln	Ser	Gly	Glu	Thr	Gly
	225				230						235				240	
30	Glu	Val	Gln	Ala	Cys	Pro	Asp	Lys	Arg	Tyr	Ala	Met	Val	Leu	Phe	Arg
				245						250					255	
	Ile	Thr	Ser	Val	Phe	Tyr	Ile	Leu	Trp	Leu	Pro	Tyr	Ile	Ile	Tyr	Phe
				260					265					270		
35	Leu	Leu	Glu	Ser	Ser	Thr	Gly	His	Ser	Asn	Arg	Phe	Ala	Ser	Phe	Leu
			275					280					285			
	Thr	Thr	Trp	Leu	Ala	Ile	Ser	Asn	Ser	Phe	Cys	Asn	Cys	Val	Ile	Tyr
		290					295					300				

Ser Leu Ser Asn Ser Val Phe Gln Arg Gly Leu Lys Arg Leu Ser Gly
305 310 315 320

Ala Met Cys Thr Ser Cys Ala Ser Gln Thr Thr Ala Asn Asp Pro Tyr
325 330 335

5 Thr Val Arg Ser Lys Gly Pro Leu Asn Gly Cys His Ile
340 345

(46) INFORMATION FOR SEQ ID NO:45:

- 10 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: DNA (genomic)
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

15 TCCCCCGGGA AAAAAACCAA CTGCTCCAAA 30

(47) INFORMATION FOR SEQ ID NO:46:

- 20 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 31 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: DNA (genomic)
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

TAGGATCCAT TTGAATGTGG ATTTGGTGAA A 31

25 (48) INFORMATION FOR SEQ ID NO:47:

- 30 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1302 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: DNA (genomic)
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

ATGTGTTTTT CTCCATTCT GGAAATCAAC ATGCAGTCTG AATCTAACAT TACAGTGCGA 60

GATGACATTG ATGACATCAA CACCAATATG TACCAACCAC TATCATATCC GTTAAGCTTT 120

35 CAAGTGCTCTC TCACCGGATT TCTTATGTGA GAAATTGTGT TGGGACTTGG CAGCAACCTC 180

ACTGTATTGG TACTTTACTG CATGAAATCC AACTTAATCA ACTCTGTCAG TAACATTATT 240
 ACAATGAATC TTCATGTACT TGATGTAATA ATTTGTGTGG GATGTATTCC TCTAACTATA 300
 GTTATCCTTC TGCTTTCACT GGAGAGTAAC ACTGCTCTCA TTTGCTGTTT CCATGAGGCT 360
 TGTGTATCTT TTGCAAGTGT CTCAACAGCA ATCAACGTTT TTGCTATCAC TTTGGACAGA 420
 5 TATGACATCT CTGTAAAACC TGCAAACCGA ATTCTGACAA TGGGCAGAGC TGTAATGTTA 480
 ATGATATCCA TTTGGATTTT TTCTTTTTTC TCTTTCCTGA TTCCTTTTAT TGAGGTAAAT 540
 TTTTTCAGTC TTCAAAGTGG AAATACCTGG GAAAACAAGA CACTTTTATG TGTCAGTACA 600
 AATGAATACT AACTGAACT GGGAATGTAT TATCACCTGT TAGTACAGAT CCCAATATTC 660
 TTTTTCAGTG TTGTAGTAAT GTTAATCACA TACACCAAAA TACTTCAGGC TCTTAATATT 720
 10 CGAATAGGCA CAAGATTTTC AACAGGGCAG AAGAAGAAAG CAAGAAAGAA AAAGACAATT 780
 TCTCTAACCA CACAACATGA GGCTACAGAC ATGTCACAAA GCAGTGGTGG GAGAAATGTA 840
 GTCTTTGGTG TAAGAACTTC AGTTTCTGTA ATAATTGCCC TCCGGCGAGC TGTGAAACGA 900
 CACCGTGAAC GACGAGAAAG ACAAAGAGA GTCTTCAGGA TGTCTTTATT GATTATTTCT 960
 ACATTTCTTC TCTGCTGGAC ACCAATTCTT GTTTTAAATA CCACCATTTT ATGTTTAGGC 1020
 15 CCAAGTGACC TTTTAGTAAA ATTAAGATTG TGTTTTTTAG TCATGGCTTA TGAACAACCT 1080
 ATATTTCACT CTCTATTATA TGCATTCCT AGACAAAAAT TTCAAAGGT CTTGAAAAGT 1140
 AAAATGAAAA AGCGAGTTGT TTCTATAGTA GAAGCTGATC CCCTGCCTAA TAATGCTGTA 1200
 ATACACAACCT CTTGGATAGA TCCCAAAGA AACAAAAAA TTACCTTTGA AGATAGTGAA 1260
 ATAAGAGAAA AACGTTTAGT GCCTCAGGTT GTCACAGACT AG 1302

20 (49) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 433 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

25 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

Met Cys Phe Ser Pro Ile Leu Glu Ile Asn Met Gln Ser Glu Ser Asn
 1 5 10 15

30 Ile Thr Val Arg Asp Asp Ile Asp Asp Ile Asn Thr Asn Met Tyr Gln
 20 25 30

Pro Leu Ser Tyr Pro Leu Ser Phe Gln Val Ser Leu Thr Gly Phe Leu
 35 40 45
 Met Leu Glu Ile Val Leu Gly Leu Gly Ser Asn Leu Thr Val Leu Val
 50 55 60
 5 Leu Tyr Cys Met Lys Ser Asn Leu Ile Asn Ser Val Ser Asn Ile Ile
 65 70 75 80
 Thr Met Asn Leu His Val Leu Asp Val Ile Ile Cys Val Gly Cys Ile
 85 90 95
 10 Pro Leu Thr Ile Val Ile Leu Leu Leu Ser Leu Glu Ser Asn Thr Ala
 100 105 110
 Leu Ile Cys Cys Phe His Glu Ala Cys Val Ser Phe Ala Ser Val Ser
 115 120 125
 Thr Ala Ile Asn Val Phe Ala Ile Thr Leu Asp Arg Tyr Asp Ile Ser
 130 135 140
 15 Val Lys Pro Ala Asn Arg Ile Leu Thr Met Gly Arg Ala Val Met Leu
 145 150 155 160
 Met Ile Ser Ile Trp Ile Phe Ser Phe Phe Ser Phe Leu Ile Pro Phe
 165 170 175
 20 Ile Glu Val Asn Phe Phe Ser Leu Gln Ser Gly Asn Thr Trp Glu Asn
 180 185 190
 Lys Thr Leu Leu Cys Val Ser Thr Asn Glu Tyr Tyr Thr Glu Leu Gly
 195 200 205
 Met Tyr Tyr His Leu Leu Val Gln Ile Pro Ile Phe Phe Phe Thr Val
 210 215 220
 25 Val Val Met Leu Ile Thr Tyr Thr Lys Ile Leu Gln Ala Leu Asn Ile
 225 230 235 240
 Arg Ile Gly Thr Arg Phe Ser Thr Gly Gln Lys Lys Lys Ala Arg Lys
 245 250 255
 30 Lys Lys Thr Ile Ser Leu Thr Thr Gln His Glu Ala Thr Asp Met Ser
 260 265 270
 Gln Ser Ser Gly Gly Arg Asn Val Val Phe Gly Val Arg Thr Ser Val
 275 280 285
 Ser Val Ile Ile Ala Leu Arg Arg Ala Val Lys Arg His Arg Glu Arg
 290 295 300
 35 Arg Glu Arg Gln Lys Arg Val Phe Arg Met Ser Leu Leu Ile Ile Ser
 305 310 315 320
 Thr Phe Leu Leu Cys Trp Thr Pro Ile Ser Val Leu Asn Thr Thr Ile

325 330 335

Leu Val Met Ala Tyr Gly Thr Thr Ile Phe His Pro Leu Leu Tyr Ala
5 355 360 365

Arg Val Val Ser Ile Val Glu Ala Asp Pro Leu Pro Asn Asn Ala Val
385 390 395 400

Glu Asp Ser Glu Ile Arg Glu Lys Arg Leu Val Pro Gln Val Val Thr
420 425 430

15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

GTGAAGCTTG CCTCTGGTGC CTGCAGGAGG

30

25 (51) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

GCAGAATTCC CGGTGGCGTG TTGTGGTGCC C

31

(52) INFORMATION FOR SEQ ID NO:51:

35 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1209 base pairs

(B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

	ATGTTGTGTC CTTCCAAGAC AGATGGCTCA GGGCACTCTG GTAGGATTCA CCAGGAAACT	60
	CATGGAGAAG GGAAAAGGGA CAAGATTAGC AACAGTGAAG GGAGGGAGAA TGGTGGGAGA	120
	GGATTCCAGA TGAACGGTGG GTCGCTGGAG GCTGAGCATG CCAGCAGGAT GTCAGTTCTC	180
	AGAGCAAAGC CCATGTCAA CAGCCAACGC TTGCTCCTTC TGTCCCCAGG ATCACCTCCT	240
10	CGCACGGGGA GCATCTCCTA CATCAACATC ATCATGCCTT CGGTGTTTCGG CACCATCTGC	300
	CTCCTGGGCA TCATCGGGAA CTCCACGGTC ATCTTCGCGG TCGTGAAGAA GTCCAAGCTG	360
	CACTGGTGCA ACAACGTCCC CGACATCTTC ATCATCAACC TCTCGGTAGT AGATCTCCTC	420
	TTTCTCCTGG GCATGCCCTT CATGATCCAC CAGCTCATGG GCAATGGGGT GTGGCACTTT	480
	GGGGAGACCA TGTGCACCCT CATCACGGCC ATGGATGCCA ATAGTCAGTT CACCAGCACC	540
15	TACATCCTGA CCGCCATGGC CATTGACCGC TACCTGGCCA CTGTCCACCC CATCTCTTCC	600
	ACGAAGTTCC GGAAGCCCTC TGTGGCCACC CTGGTGATCT GCCTCCTGTG GGCCCTCTCC	660
	TTCATCAGCA TCACCCCTGT GTGGCTGTAT GCCAGACTCA TCCCCTTCCC AGGAGGTGCA	720
	GTGGGCTGCG GCATACGCCT GCCCAACCCA GACACTGACC TCTACTGGTT CACCCTGTAC	780
	CAGTTTTTCC TGGCCTTTGC CCTGCCTTTT GTGGTCATCA CAGCCGCATA CGTGAGGATC	840
20	CTGCAGCGCA TGACGTCCTC AGTGGCCCCC GCCTCCCAGC GCAGCATCCG GCTGCGGACA	900
	AAGAGGGTGA CCCGCACAGC CATCGCCATC TGTCTGGTCT TCTTTGTGTG CTGGGCACCC	960
	TACTATGTGC TACAGCTGAC CCAGTTGTCC ATCAGCCGCC CGACCCTCAC CTTTGTCTAC	1020
	TTATACAATG CGGCCATCAG CTTGGGCTAT GCCAACAGCT GCCTCAACCC CTTTGTGTAC	1080
	ATCGTGCTCT GTGAGACGTT CCGCAAACGC TTGGTCCTGT CGGTGAAGCC TGCAGCCCAG	1140
25	GGGCAGCTTC GCGCTGTCAG CAACGCTCAG ACGGCTGACG AGGAGAGGAC AGAAAGCAAA	1200
	GGCACCTGA	1209

(53) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 402 amino acids
 (B) TYPE: amino acid

30

41

(C) STRANDEDNESS:

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

5	Met	Leu	Cys	Pro	Ser	Lys	Thr	Asp	Gly	Ser	Gly	His	Ser	Gly	Arg	Ile	1	5	10	15
	His	Gln	Glu	Thr	His	Gly	Glu	Gly	Lys	Arg	Asp	Lys	Ile	Ser	Asn	Ser	20	25	30	
10	Glu	Gly	Arg	Glu	Asn	Gly	Gly	Arg	Gly	Phe	Gln	Met	Asn	Gly	Gly	Ser	35	40	45	
	Leu	Glu	Ala	Glu	His	Ala	Ser	Arg	Met	Ser	Val	Leu	Arg	Ala	Lys	Pro	50	55	60	
	Met	Ser	Asn	Ser	Gln	Arg	Leu	Leu	Leu	Ser	Pro	Gly	Ser	Pro	Pro		65	70	75	80
15	Arg	Thr	Gly	Ser	Ile	Ser	Tyr	Ile	Asn	Ile	Ile	Met	Pro	Ser	Val	Phe	85	90	95	
	Gly	Thr	Ile	Cys	Leu	Leu	Gly	Ile	Ile	Gly	Asn	Ser	Thr	Val	Ile	Phe	100	105	110	
20	Ala	Val	Val	Lys	Lys	Ser	Lys	Leu	His	Trp	Cys	Asn	Asn	Val	Pro	Asp	115	120	125	
	Ile	Phe	Ile	Ile	Asn	Leu	Ser	Val	Val	Asp	Leu	Leu	Phe	Leu	Leu	Gly	130	135	140	
	Met	Pro	Phe	Met	Ile	His	Gln	Leu	Met	Gly	Asn	Gly	Val	Trp	His	Phe	145	150	155	160
25	Gly	Glu	Thr	Met	Cys	Thr	Leu	Ile	Thr	Ala	Met	Asp	Ala	Asn	Ser	Gln	165	170	175	
	Phe	Thr	Ser	Thr	Tyr	Ile	Leu	Thr	Ala	Met	Ala	Ile	Asp	Arg	Tyr	Leu	180	185	190	
30	Ala	Thr	Val	His	Pro	Ile	Ser	Ser	Thr	Lys	Phe	Arg	Lys	Pro	Ser	Val	195	200	205	
	Ala	Thr	Leu	Val	Ile	Cys	Leu	Leu	Trp	Ala	Leu	Ser	Phe	Ile	Ser	Ile	210	215	220	
	Thr	Pro	Val	Trp	Leu	Tyr	Ala	Arg	Leu	Ile	Pro	Phe	Pro	Gly	Gly	Ala	225	230	235	240
35	Val	Gly	Cys	Gly	Ile	Arg	Leu	Pro	Asn	Pro	Asp	Thr	Asp	Leu	Tyr	Trp	245	250	255	

Phe Thr Leu Tyr Gln Phe Phe Leu Ala Phe Ala Leu Pro Phe Val Val
260 265 270

Ile Thr Ala Ala Tyr Val Arg Ile Leu Gln Arg Met Thr Ser Ser Val
275 280 285

5 Ala Pro Ala Ser Gln Arg Ser Ile Arg Leu Arg Thr Lys Arg Val Thr
290 295 300

Arg Thr Ala Ile Ala Ile Cys Leu Val Phe Phe Val Cys Trp Ala Pro
305 310 315 320

10 Tyr Tyr Val Leu Gln Leu Thr Gln Leu Ser Ile Ser Arg Pro Thr Leu
325 330 335

Thr Phe Val Tyr Leu Tyr Asn Ala Ala Ile Ser Leu Gly Tyr Ala Asn
340 345 350

Ser Cys Leu Asn Pro Phe Val Tyr Ile Val Leu Cys Glu Thr Phe Arg
355 360 365

15 Lys Arg Leu Val Leu Ser Val Lys Pro Ala Ala Gln Gly Gln Leu Arg
 370 375 380

Ala Val Ser Asn Ala Gln Thr Ala Asp Glu Glu Arg Thr Glu Ser Lys
385 390 395 400

Gly Thr

(54) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

GGCGGATCCA TGGATGTGAC TTCCCAA

27

30 (55) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

GGCGGATCCC TACACGGCAC TGCTGAA

27

(56) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1128 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

10 ATGGATGTGA CTTCCCAAGC CCGGGGCGTG GGCCTGGAGA TGTACCCAGG CACCGCGCAC 60
GCTGCGGCCC CCAACACCAC CTCCCCGAG CTCAACCTGT CCCACCCGCT CCTGGGCACC 120
GCCCTGGCCA ATGGGACAGG TGAGCTCTCG GAGCACCAGC AGTACGTGAT CGGCCTGTTC 180
CTCTCGTGCC TCTACACCAT CTTCTCTTC CCCATCGGCT TTGTGGGCAA CATCCTGATC 240
CTGGTGGTGA ACATCAGCTT CCGCGAGAAG ATGACCATCC CCGACCTGTA CTTTCATCAAC 300
15 CTGGCGGTGG CGGACCTCAT CCTGGTGGCC GACTCCCTCA TTGAGGTGTT CAACCTGCAC 360
GAGCGGTACT ACGACATCGC CGTCCTGTGC ACCTTCATGT CGCTCTTCCT GCAGGTCAAC 420
ATGTACAGCA GCGTCTTCTT CCTCACCTGG ATGAGCTTCG ACCGCTACAT CGCCCTGGCC 480
AGGGCCATGC GCTGCAGCCT GTTCCGCACC AAGCACCACG CCCGGCTGAG CTGTGGCCTC 540
ATCTGGATGG CATCCGTGTC AGCCACGCTG GTGCCCTTCA CCGCCGTGCA CCTGCAGCAC 600
20 ACCGACGAGG CCTGCTTCTG TTTCGCGGAT GTCCGGGAGG TGCAGTGGCT CGAGGTCACG 660
CTGGGCTTCA TCGTGCCCTT CGCCATCATC GGCCTGTGCT ACTCCCTCAT TGTCCGGGTG 720
CTGGTCAGGG CGCACC GGCA CCGTGGGCTG CGGCCCCGGC GGCAGAAGGC GCTCCGCATG 780
ATCCTCGCGG TGGTGCTGGT CTTCTTCGTC TGCTGGCTGC CGGAGAACGT CTTTCATCAGC 840
GTGCACCTCC TGCAGCGGAC GCAGCCTGGG GCCGCTCCCT GCAAGCAGTC TTTCCGCCAT 900
25 GCCCACCACC TCACGGGCCA CATTGTCAAC CTCGCCGCCT TCTCCAACAG CTGCCTAAAC 960
CCCCTCATCT ACAGCTTTCT CGGGGAGACC TTCAGGGACA AGCTGAGGCT GTACATTGAG 1020
CAGAAAACAA ATTTGCCGGC CCTGAACCGC TTCTGTACG CTGCCCTGAA GGCCGTCATT 1080
CCAGACAGCA CCGAGCAGTC GGATGTGAGG TTCAGCAGTG CCGTGTGA 1128

(57) INFORMATION FOR SEQ ID NO:56:

30

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 375 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: not relevant

5 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

	Met	Asp	Val	Thr	Ser	Gln	Ala	Arg	Gly	Val	Gly	Leu	Glu	Met	Tyr	Pro	
	1				5					10						15	
10	Gly	Thr	Ala	His	Ala	Ala	Ala	Pro	Asn	Thr	Thr	Ser	Pro	Glu	Leu	Asn	
				20					25					30			
	Leu	Ser	His	Pro	Leu	Leu	Gly	Thr	Ala	Leu	Ala	Asn	Gly	Thr	Gly	Glu	
			35					40					45				
	Leu	Ser	Glu	His	Gln	Gln	Tyr	Val	Ile	Gly	Leu	Phe	Leu	Ser	Cys	Leu	
		50					55					60					
15	Tyr	Thr	Ile	Phe	Leu	Phe	Pro	Ile	Gly	Phe	Val	Gly	Asn	Ile	Leu	Ile	
	65					70					75					80	
	Leu	Val	Val	Asn	Ile	Ser	Phe	Arg	Glu	Lys	Met	Thr	Ile	Pro	Asp	Leu	
				85						90					95		
20	Tyr	Phe	Ile	Asn	Leu	Ala	Val	Ala	Asp	Leu	Ile	Leu	Val	Ala	Asp	Ser	
				100					105						110		
	Leu	Ile	Glu	Val	Phe	Asn	Leu	His	Glu	Arg	Tyr	Tyr	Asp	Ile	Ala	Val	
			115					120					125				
	Leu	Cys	Thr	Phe	Met	Ser	Leu	Phe	Leu	Gln	Val	Asn	Met	Tyr	Ser	Ser	
		130					135					140					
25	Val	Phe	Phe	Leu	Thr	Trp	Met	Ser	Phe	Asp	Arg	Tyr	Ile	Ala	Leu	Ala	
	145					150					155					160	
	Arg	Ala	Met	Arg	Cys	Ser	Leu	Phe	Arg	Thr	Lys	His	His	Ala	Arg	Leu	
				165						170						175	
30	Ser	Cys	Gly	Leu	Ile	Trp	Met	Ala	Ser	Val	Ser	Ala	Thr	Leu	Val	Pro	
			180						185					190			
	Phe	Thr	Ala	Val	His	Leu	Gln	His	Thr	Asp	Glu	Ala	Cys	Phe	Cys	Phe	
			195					200					205				
	Ala	Asp	Val	Arg	Glu	Val	Gln	Trp	Leu	Glu	Val	Thr	Leu	Gly	Phe	Ile	
		210					215					220					
35	Val	Pro	Phe	Ala	Ile	Ile	Gly	Leu	Cys	Tyr	Ser	Leu	Ile	Val	Arg	Val	
	225					230					235					240	
	Leu	Val	Arg	Ala	His	Arg	His	Arg	Gly	Leu	Arg	Pro	Arg	Arg	Gln	Lys	

45

		245		250		255										
	Ala	Leu	Arg	Met	Ile	Leu	Ala	Val	Val	Leu	Val	Phe	Phe	Val	Cys	Trp
				260					265					270		
5	Leu	Pro	Glu	Asn	Val	Phe	Ile	Ser	Val	His	Leu	Leu	Gln	Arg	Thr	Gln
			275					280					285			
	Pro	Gly	Ala	Ala	Pro	Cys	Lys	Gln	Ser	Phe	Arg	His	Ala	His	Pro	Leu
		290					295				300					
	Thr	Gly	His	Ile	Val	Asn	Leu	Ala	Ala	Phe	Ser	Asn	Ser	Cys	Leu	Asn
	305				310					315					320	
10	Pro	Leu	Ile	Tyr	Ser	Phe	Leu	Gly	Glu	Thr	Phe	Arg	Asp	Lys	Leu	Arg
				325						330					335	
	Leu	Tyr	Ile	Glu	Gln	Lys	Thr	Asn	Leu	Pro	Ala	Leu	Asn	Arg	Phe	Cys
			340					345					350			
15	His	Ala	Ala	Leu	Lys	Ala	Val	Ile	Pro	Asp	Ser	Thr	Glu	Gln	Ser	Asp
		355					360						365			
	Val	Arg	Phe	Ser	Ser	Ala	Val									
		370				375										

(58) INFORMATION FOR SEQ ID NO:57:

- 20 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 31 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

- 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

AAGGAATTCA CGGCCGGGTG ATGCCATTCC C

31

(59) INFORMATION FOR SEQ ID NO:58:

- 30 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

35 GGTGGATCCA TAAACACGGG CGTTGAGGAC

30

(60) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 960 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

ATGCCATTCC CAAACTGCTC AGCCCCCAGC ACTGTGGTGG CCACAGCTGT GGGTGTCTTG 60
 CTGGGGCTGG AGTGTGGGCT GGGTCTGCTG GGCAACGCGG TGGCGCTGTG GACCTTCCTG 120
 10 TTCCGGGTCA GGGTGTGGAA GCCGTACGCT GTCTACCTGC TCAACCTGGC CCTGGCTGAC 180
 CTGCTGTTGG CTGCGTGCCT GCCTTTCCTG GCCGCCTTCT ACCTGAGCCT CCAGGCTTGG 240
 CATCTGGGCC GTGTGGGCTG CTGGGCCCTG CGCTTCCTGC TGGACCTCAG CCGCAGCGTG 300
 GGGATGGCCT TCCTGGCCGC CGTGGCTTTG GACCGGTACC TCCGTGTGGT CCACCCTCGG 360
 CTTAAGGTCA ACCTGCTGTC TCCTCAGGCG GCCCTGGGGG TCTCGGGCCT CGTCTGGCTC 420
 15 CTGATGGTCG CCCTCACCTG CCCGGGCTTG CTCATCTCTG AGGCCGCCCA GAACTCCACC 480
 AGGTGCCACA GTTCTACTC CAGGGCAGAC GGCTCCTTCA GCATCATCTG GCAGGAAGCA 540
 CTCTCCTGCC TTCAGTTTGT CCTCCCCTTT GGCCTCATCG TGTTCTGCAA TGCAGGCATC 600
 ATCAGGGCTC TCCAGAAAAG ACTCCGGGAG CCTGAGAAAC AGCCCAAGCT TCAGCGGGCC 660
 CAGGCACTGG TCACCTTGGT GGTGGTGCTG TTTGCTCTGT GCTTCTGCC CTGCTTCCTG 720
 20 GCCAGAGTCC TGATGCACAT CTTCCAGAAT CTGGGGAGCT GCAGGGCCCT TTGTGCAGTG 780
 GCTCATACCT CGGATGTCAC GGGCAGCCTC ACCTACCTGC ACAGTGTCTG CAACCCCGTG 840
 GTATACTGCT TCTCCAGCCC CACCTTCAGG AGCTCCTATC GGAGGGTCTT CCACACCCTC 900
 CGAGGCAAAG GGCAGGCAGC AGAGCCCCCA GATTTC AACC CCAGAGACTC CTATTCCTGA 960

(61) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 319 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

Met Pro Phe Pro Asn Cys Ser Ala Pro Ser Thr Val Val Ala Thr Ala

47

	1		5			10		15								
	Val	Gly	Val	Leu	Leu	Gly	Leu	Glu	Cys	Gly	Leu	Gly	Leu	Leu	Gly	Asn
				20				25					30			
5	Ala	Val	Ala	Leu	Trp	Thr	Phe	Leu	Phe	Arg	Val	Arg	Val	Trp	Lys	Pro
			35				40						45			
	Tyr	Ala	Val	Tyr	Leu	Leu	Asn	Leu	Ala	Leu	Ala	Asp	Leu	Leu	Leu	Ala
		50					55					60				
	Ala	Cys	Leu	Pro	Phe	Leu	Ala	Ala	Phe	Tyr	Leu	Ser	Leu	Gln	Ala	Trp
	65					70					75				80	
10	His	Leu	Gly	Arg	Val	Gly	Cys	Trp	Ala	Leu	Arg	Phe	Leu	Leu	Asp	Leu
					85					90					95	
	Ser	Arg	Ser	Val	Gly	Met	Ala	Phe	Leu	Ala	Ala	Val	Ala	Leu	Asp	Arg
				100					105					110		
15	Tyr	Leu	Arg	Val	Val	His	Pro	Arg	Leu	Lys	Val	Asn	Leu	Leu	Ser	Pro
			115					120					125			
	Gln	Ala	Ala	Leu	Gly	Val	Ser	Gly	Leu	Val	Trp	Leu	Leu	Met	Val	Ala
		130					135					140				
	Leu	Thr	Cys	Pro	Gly	Leu	Leu	Ile	Ser	Glu	Ala	Ala	Gln	Asn	Ser	Thr
	145					150				155						160
20	Arg	Cys	His	Ser	Phe	Tyr	Ser	Arg	Ala	Asp	Gly	Ser	Phe	Ser	Ile	Ile
					165					170					175	
	Trp	Gln	Glu	Ala	Leu	Ser	Cys	Leu	Gln	Phe	Val	Leu	Pro	Phe	Gly	Leu
				180					185					190		
25	Ile	Val	Phe	Cys	Asn	Ala	Gly	Ile	Ile	Arg	Ala	Leu	Gln	Lys	Arg	Leu
			195					200					205			
	Arg	Glu	Pro	Glu	Lys	Gln	Pro	Lys	Leu	Gln	Arg	Ala	Gln	Ala	Leu	Val
		210					215					220				
	Thr	Leu	Val	Val	Val	Leu	Phe	Ala	Leu	Cys	Phe	Leu	Pro	Cys	Phe	Leu
	225					230					235					240
30	Ala	Arg	Val	Leu	Met	His	Ile	Phe	Gln	Asn	Leu	Gly	Ser	Cys	Arg	Ala
					245					250					255	
	Leu	Cys	Ala	Val	Ala	His	Thr	Ser	Asp	Val	Thr	Gly	Ser	Leu	Thr	Tyr
				260					265					270		
35	Leu	His	Ser	Val	Val	Asn	Pro	Val	Val	Tyr	Cys	Phe	Ser	Ser	Pro	Thr
			275					280					285			
	Phe	Arg	Ser	Ser	Tyr	Arg	Arg	Val	Phe	His	Thr	Leu	Arg	Gly	Lys	Gly
		290					295						300			

• 20

315

5

(ii) MOLECULE TYPE: DNA (genomic)

10

20

25

30

(63) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 380 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

5
 10
 15
 20
 25
 30
 35

Met Glu Glu Gly Gly Asp Phe Asp Asn Tyr Tyr Gly Ala Asp Asn Gln
 1 5 10 15
 Ser Glu Cys Glu Tyr Thr Asp Trp Lys Ser Ser Gly Ala Leu Ile Pro
 20 25 30
 Ala Ile Tyr Met Leu Val Phe Leu Leu Gly Thr Thr Gly Asn Gly Leu
 35 40 45
 Val Leu Trp Thr Val Phe Arg Ser Ser Arg Glu Lys Arg Arg Ser Ala
 50 55 60
 Asp Ile Phe Ile Ala Ser Leu Ala Val Ala Asp Leu Thr Phe Val Val
 65 70 75 80
 Thr Leu Pro Leu Trp Ala Thr Tyr Thr Tyr Arg Asp Tyr Asp Trp Pro
 85 90 95
 Phe Gly Thr Phe Phe Cys Lys Leu Ser Ser Tyr Leu Ile Phe Val Asn
 100 105 110
 Met Tyr Ala Ser Val Phe Cys Leu Thr Gly Leu Ser Phe Asp Arg Tyr
 115 120 125
 Leu Ala Ile Val Arg Pro Val Ala Asn Ala Arg Leu Arg Leu Arg Val
 130 135 140
 Ser Gly Ala Val Ala Thr Ala Val Leu Trp Val Leu Ala Ala Leu Leu
 145 150 155 160
 Ala Met Pro Val Met Val Leu Arg Thr Thr Gly Asp Leu Glu Asn Thr
 165 170 175
 Thr Lys Val Gln Cys Tyr Met Asp Tyr Ser Met Val Ala Thr Val Ser
 180 185 190
 Ser Glu Trp Ala Trp Glu Val Gly Leu Gly Val Ser Ser Thr Thr Val
 195 200 205
 Gly Phe Val Val Pro Phe Thr Ile Met Leu Thr Cys Tyr Phe Phe Ile
 210 215 220
 Ala Gln Thr Ile Ala Gly His Phe Arg Lys Glu Arg Ile Glu Gly Leu

225		230		235		240
Arg Lys Arg Arg Arg Leu Leu Ser Ile Ile Val Val Leu Val Val Thr						
	245		250		255	
5 Phe Ala Leu Cys Trp Met Pro Tyr His Leu Val Lys Thr Leu Tyr Met	260		265		270	
Leu Gly Ser Leu Leu His Trp Pro Cys Asp Phe Asp Leu Phe Leu Met	275		280		285	
Asn Ile Phe Pro Tyr Cys Thr Cys Ile Ser Tyr Val Asn Ser Cys Leu	290		295		300	
10 Asn Pro Phe Leu Tyr Ala Phe Phe Asp Pro Arg Phe Arg Gln Ala Cys	305		310		315	320
Thr Ser Met Leu Cys Cys Gly Gln Ser Arg Cys Ala Gly Thr Ser His		325		330		335
15 Ser Ser Ser Gly Glu Lys Ser Ala Ser Tyr Ser Ser Gly His Ser Gln	340		345		350	
Gly Pro Gly Pro Asn Met Gly Lys Gly Gly Glu Gln Met His Glu Lys	355		360		365	
Ser Ile Pro Tyr Ser Gln Glu Thr Leu Val Val Asp	370		375		380	

20 (64) INFORMATION FOR SEQ ID NO:63:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 31 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- 25
- (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

TGAGAATTCT GGTGACTCAC AGCCGGCACA G

31

(65) INFORMATION FOR SEQ ID NO:64:

- 30 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 31 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- 35
- (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

GCCGGATCCA AGGAAAAGCA GCAATAAAAG G

31

(66) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1119 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

10 ATGAACTACC CGCTAACGCT GGAAATGGAC CTCGAGAACC TGGAGGACCT GTTCTGGGAA 60
 CTGGACAGAT TGGACAAC TAACGACACC TCCCTGGTGG AAAATCATCT CTGCCCTGCC 120
 ACAGAGGGTC CCCTCATGGC CTCCTTCAAG GCCGTGTTCG TGCCCGTGGC CTACAGCCTC 180
 ATCTTCCTCC TGGGCGTGAT CGGCAACGTC CTGGTGCTGG TGATCCTGGA GCGGCACCGG 240
 CAGACACGCA GTTCCACGGA GACCTTCCTG TTCCACCTGG CCGTGGCCGA CCTCCTGCTG 300
 15 GTCTTCATCT TGCCCTTTGC CGTGGCCGAG GGCTCTGTGG GCTGGGTCCT GGGGACCTTC 360
 CTCTGCAAAA CTGTGATTGC CCTGCACAAA GTCAACTTCT ACTGCAGCAG CCTGCTCCTG 420
 GCCTGCATCG CCGTGGACCG CTACCTGGCC ATTGTCCACG CCGTCCATGC CTACCGCCAC 480
 CGCCGCTCC TCTCCATCCA CATCACCTGT GGGACCATCT GGCTGGTGGG CTTCTCCTT 540
 GCCTTGCCAG AGATTCTCTT CGCCAAAGTC AGCCAAGGCC ATCACAACAA CTCCCTGCCA 600
 20 CGTTGCACCT TCTCCCAAGA GAACCAAGCA GAAACGCATG CCTGGTTCAC CTCCCGATTC 660
 CTCTACCATG TGGCGGGATT CCTGCTGCCC ATGCTGGTGA TGGGCTGGTG CTACGTGGGG 720
 GTAGTGACACA GGTTCGCGCA GGCCCAGCGG CGCCCTCAGC GGCAGAAGGC AGTCAGGGTG 780
 GCCATCCTGG TGACAAGCAT CTTCTTCCTC TGCTGGTCAC CCTACCACAT CGTCATCTTC 840
 CTGGACACCC TGGCGAGGCT GAAGGCCGTG GACAATACCT GCAAGCTGAA TGGCTCTCTC 900
 25 CCCGTGGCCA TCACCATGTG TGAGTTCCTG GGCCTGGCCC ACTGCTGCCT CAACCCCATG 960
 CTCTACACTT TCGCCGGCGT GAAGTTCCGC AGTGACCTGT CGCGGCTCCT GACCAAGCTG 1020
 GGCTGTACCG GCCCTGCCTC CCTGTGCCAG CTCTTCCTA GCTGGCGCAG GAGCAGTCTC 1080
 TCTGAGTCAG AGAATGCCAC CTCTCTCACC ACGTTCTAG 1119

(67) INFORMATION FOR SEQ ID NO:66:

30

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 372 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

5 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

```

Met Asn Tyr Pro Leu Thr Leu Glu Met Asp Leu Glu Asn Leu Glu Asp
1              5              10              15

Leu Phe Trp Glu Leu Asp Arg Leu Asp Asn Tyr Asn Asp Thr Ser Leu
10              20              25              30

Val Glu Asn His Leu Cys Pro Ala Thr Glu Gly Pro Leu Met Ala Ser
              35              40              45

Phe Lys Ala Val Phe Val Pro Val Ala Tyr Ser Leu Ile Phe Leu Leu
              50              55              60

15 Gly Val Ile Gly Asn Val Leu Val Leu Val Ile Leu Glu Arg His Arg
65              70              75              80

Gln Thr Arg Ser Ser Thr Glu Thr Phe Leu Phe His Leu Ala Val Ala
              85              90              95

20 Asp Leu Leu Leu Val Phe Ile Leu Pro Phe Ala Val Ala Glu Gly Ser
              100             105             110

Val Gly Trp Val Leu Gly Thr Phe Leu Cys Lys Thr Val Ile Ala Leu
              115             120             125

His Lys Val Asn Phe Tyr Cys Ser Ser Leu Leu Leu Ala Cys Ile Ala
              130             135             140

25 Val Asp Arg Tyr Leu Ala Ile Val His Ala Val His Ala Tyr Arg His
145             150             155             160

Arg Arg Leu Leu Ser Ile His Ile Thr Cys Gly Thr Ile Trp Leu Val
              165             170             175

30 Gly Phe Leu Leu Ala Leu Pro Glu Ile Leu Phe Ala Lys Val Ser Gln
              180             185             190

Gly His His Asn Asn Ser Leu Pro Arg Cys Thr Phe Ser Gln Glu Asn
              195             200             205

Gln Ala Glu Thr His Ala Trp Phe Thr Ser Arg Phe Leu Tyr His Val
              210             215             220

35 Ala Gly Phe Leu Leu Pro Met Leu Val Met Gly Trp Cys Tyr Val Gly
225             230             235             240

Val Val His Arg Leu Arg Gln Ala Gln Arg Arg Pro Gln Arg Gln Lys

```

53

		245		250		255
	Ala Val Arg Val Ala Ile Leu Val Thr Ser Ile Phe Phe Leu Cys Trp					
		260		265		270
5	Ser Pro Tyr His Ile Val Ile Phe Leu Asp Thr Leu Ala Arg Leu Lys					
		275		280		285
	Ala Val Asp Asn Thr Cys Lys Leu Asn Gly Ser Leu Pro Val Ala Ile					
		290		295		300
	Thr Met Cys Glu Phe Leu Gly Leu Ala His Cys Cys Leu Asn Pro Met					
		305		310		315
10	Leu Tyr Thr Phe Ala Gly Val Lys Phe Arg Ser Asp Leu Ser Arg Leu					
				325		330
	Leu Thr Lys Leu Gly Cys Thr Gly Pro Ala Ser Leu Cys Gln Leu Phe					
		340		345		350
15	Pro Ser Trp Arg Arg Ser Ser Leu Ser Glu Ser Glu Asn Ala Thr Ser					
		355		360		365
	Leu Thr Thr Phe					
		370				

(68) INFORMATION FOR SEQ ID NO:67:

- 20 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

CAAAGCTTGA AAGCTGCACG GTGCAGAGAC

30

(69) INFORMATION FOR SEQ ID NO:68:

- 30 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

35 GCGGATCCCC AGTCACACCC TGGCTGGGCC

30

(70) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1128 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

ATGGATGTGA CTTCCAAGC CCGGGGCGTG GGCCTGGAGA TGTACCCAGG CACCGCGCAG 60
 CCTGCGGCCC CCAACACCAC CTCCCCGAG CTCAACCTGT CCCACCCGCT CCTGGGCACC 120
 10 GCCCTGGCCA ATGGGACAGG TGAGCTCTCG GAGCACCAGC AGTACGTGAT CGGCCTGTTC 180
 CTCTCGTGCC TCTACACCAT CTTCCTCTTC CCCATCGGCT TTGTGGGCAA CATCCTGATC 240
 CTGGTGGTGA ACATCAGCTT CCGCGAGAAG ATGACCATCC CCGACCTGTA CTTTCATCAAC 300
 CTGGCGGTGG CGGACCTCAT CCTGGTGGCC GACTCCCTCA TTGAGGTGTT CAACCTGCAC 360
 GAGCGGTACT ACGACATCGC CGTCTGTGC ACCTTCATGT CGCTCTTCCT GCAGGTCAAC 420
 15 ATGTACAGCA GCGTCTTCTT CCTCACCTGG ATGAGCTTCG ACCGCTACAT CGCCCTGGCC 480
 AGGGCCATGC GCTGCAGCCT GTTCCGCACC AAGCACCACG CCCGGCTGAG CTGTGGCCTC 540
 ATCTGGATGG CATCCGTGTC AGCCACGCTG GTGCCCTTCA CCGCCGTGCA CCTGCAGCAC 600
 ACCGACGAGG CCTGCTTCTG TTTCGCGGAT GTCCGGGAGG TGCAGTGGCT CGAGGTCACG 660
 CTGGGCTTCA TCGTGCCCTT CGCCATCATC GGCCTGTGCT ACTCCCTCAT TGTCCGGGTG 720
 20 CTGGTCAGGG CGCACC GGCA CCGTGGGCTG CGGCCCCGGC GGCAGAAGGC GCTCCGCATG 780
 ATCCTCGCGG TGGTGCTGGT CTTCTTCGTC TGCTGGCTGC CGGAGAACGT CTTTCATCAGC 840
 GTGCACCTCC TGCAGCGGAC GCAGCCTGGG GCCGCTCCCT GCAAGCAGTC TTTCCGCCAT 900
 GCCCACCCTC TCACGGGCCA CATTGTCAAC CTCACCGCCT TCTCCAACAG CTGCCTAAAC 960
 CCCCTCATCT ACAGCTTTCT CGGGGAGACC TTCAGGGACA AGCTGAGGCT GTACATTGAG 1020
 25 CAGAAAACAA ATTTGCCGGC CCTGAACCGC TTCTGTCACG CTGCCCTGAA GGCCGTCATT 1080
 CCAGACAGCA CCGAGCAGTC GGATGTGAGG TTCAGCAGTG CCGTGTAG 1128

(71) INFORMATION FOR SEQ ID NO:70:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 375 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

55

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

	Met	Asp	Val	Thr	Ser	Gln	Ala	Arg	Gly	Val	Gly	Leu	Glu	Met	Tyr	Pro	
	1				5					10					15		
5	Gly	Thr	Ala	Gln	Pro	Ala	Ala	Pro	Asn	Thr	Thr	Ser	Pro	Glu	Leu	Asn	
				20					25					30			
	Leu	Ser	His	Pro	Leu	Leu	Gly	Thr	Ala	Leu	Ala	Asn	Gly	Thr	Gly	Glu	
			35					40					45				
	Leu	Ser	Glu	His	Gln	Gln	Tyr	Val	Ile	Gly	Leu	Phe	Leu	Ser	Cys	Leu	
10		50					55					60					
	Tyr	Thr	Ile	Phe	Leu	Phe	Pro	Ile	Gly	Phe	Val	Gly	Asn	Ile	Leu	Ile	
	65					70					75				80		
	Leu	Val	Val	Asn	Ile	Ser	Phe	Arg	Glu	Lys	Met	Thr	Ile	Pro	Asp	Leu	
				85						90					95		
15	Tyr	Phe	Ile	Asn	Leu	Ala	Val	Ala	Asp	Leu	Ile	Leu	Val	Ala	Asp	Ser	
				100					105					110			
	Leu	Ile	Glu	Val	Phe	Asn	Leu	His	Glu	Arg	Tyr	Tyr	Asp	Ile	Ala	Val	
			115					120					125				
	Leu	Cys	Thr	Phe	Met	Ser	Leu	Phe	Leu	Gln	Val	Asn	Met	Tyr	Ser	Ser	
20		130					135					140					
	Val	Phe	Phe	Leu	Thr	Trp	Met	Ser	Phe	Asp	Arg	Tyr	Ile	Ala	Leu	Ala	
	145					150					155					160	
	Arg	Ala	Met	Arg	Cys	Ser	Leu	Phe	Arg	Thr	Lys	His	His	Ala	Arg	Leu	
				165						170					175		
25	Ser	Cys	Gly	Leu	Ile	Trp	Met	Ala	Ser	Val	Ser	Ala	Thr	Leu	Val	Pro	
				180					185						190		
	Phe	Thr	Ala	Val	His	Leu	Gln	His	Thr	Asp	Glu	Ala	Cys	Phe	Cys	Phe	
			195					200					205				
	Ala	Asp	Val	Arg	Glu	Val	Gln	Trp	Leu	Glu	Val	Thr	Leu	Gly	Phe	Ile	
30		210					215					220					
	Val	Pro	Phe	Ala	Ile	Ile	Gly	Leu	Cys	Tyr	Ser	Leu	Ile	Val	Arg	Val	
	225					230					235					240	
	Leu	Val	Arg	Ala	His	Arg	His	Arg	Gly	Leu	Arg	Pro	Arg	Arg	Gln	Lys	
				245						250					255		
35	Ala	Leu	Arg	Met	Ile	Leu	Ala	Val	Val	Leu	Val	Phe	Phe	Val	Cys	Trp	
				260					265						270		

Leu Pro Glu Asn Val Phe Ile Ser Val His Leu Leu Gln Arg Thr Gln
 275 280 285

Pro Gly Ala Ala Pro Cys Lys Gln Ser Phe Arg His Ala His Pro Leu
 290 295 300

5 Thr Gly His Ile Val Asn Leu Thr Ala Phe Ser Asn Ser Cys Leu Asn
 305 310 315 320

Pro Leu Ile Tyr Ser Phe Leu Gly Glu Thr Phe Arg Asp Lys Leu Arg
 325 330 335

10 Leu Tyr Ile Glu Gln Lys Thr Asn Leu Pro Ala Leu Asn Arg Phe Cys
 340 345 350

His Ala Ala Leu Lys Ala Val Ile Pro Asp Ser Thr Glu Gln Ser Asp
 355 360 365

Val Arg Phe Ser Ser Ala Val
 370 375

15 (72) INFORMATION FOR SEQ ID NO:71:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

20 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

ACAGAATTCC TGTGTGGTTT TACCGCCCAG

30

(73) INFORMATION FOR SEQ ID NO:72:

25 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

30 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

CTCGGATCCA GGCAGAAGAG TGCCTATGG

30

(74) INFORMATION FOR SEQ ID NO:73:

(i) SEQUENCE CHARACTERISTICS:

35 (A) LENGTH: 1137 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

```

    ATGGACCTGG GGAAACCAAT GAAAAGCGTG CTGGTGGTGG CTCTCCTTGT CATTTTCCAG    60
5  GTATGCCTGT GTCAAGATGA GGTACACGGAC GATTACATCG GAGACAACAC CACAGTGGAC    120
    TACACTTTGT TCGAGTCTTT GTGCTCCAAG AAGGACGTGC GGAAC TTAA AGCCTGGTTC    180
    CTCCCTATCA TGTACTCCAT CATTTGTTTC GTGGGCCTAC TGGGCAATGG GCTGGTCGTG    240
    TTGACCTATA TCTATTTCAG GAGGCTCAAG ACCATGACCG ATACCTACCT GCTCAACCTG    300
    GCGGTGGCAG ACATCCTCTT CCTCCTGACC CTTCCCTTCT GGGCCTACAG CGCGGCCAAG    360
10 TCCTGGGTCT TCGGTGTCCA CTTTGTGCAAG CTCATCTTTG CCATCTACAA GATGAGCTTC    420
    TTCAGTGGCA TGCTCCTACT TCTTTGCATC AGCATTGACC GCTACGTGGC CATCGTCCAG    480
    GCTGTCTCAG CTCACCGCCA CCGTGCCCGC GTCCTTCTCA TCAGCAAGCT GTCCTGTGTG    540
    GGCATCTGGA TACTAGCCAC AGTGCTCTCC ATCCCAGAGC TCCTGTACAG TGACCTCCAG    600
    AGGAGCAGCA GTGAGCAAGC GATGCGATGC TCTCTCATCA CAGAGCATGT GGAGGCCTTT    660
15 ATCACCATCC AGGTGGCCCA GATGGTGATC GGCTTTCTGG TCCCCCTGCT GGCCATGAGC    720
    TTCTGTTACC TTGTCATCAT CCGCACCTCG CTCCAGGCAC GCAACTTTGA GCGCAACAAG    780
    GCCATCAAGG TGATCATCGC TGTGGTCGTG GTCTTCATAG TCTTCCAGCT GCCCTACAAT    840
    GGGGTGGTCC TGGCCCAGAC GGTGGCCAAC TTCAACATCA CCAGTAGCAC CTGTGAGCTC    900
    AGTAAGCAAC TCAACATCGC CTACGACGTC ACCTACAGCC TGGCCTGCGT CCGCTGCTGC    960
20 GTCAACCCTT TCTTGTACGC CTTTCATCGG GTCAAGTTCC GCAACGATCT CTTCAAGCTC   1020
    TTCAAGGACC TGGGCTGCCT CAGCCAGGAG CAGCTCCGGC AGTGGTCTTC CTGTGGGCAC   1080
    ATCCGGCGCT CCTCCATGAG TGTGGAGGCC GAGACCACCA CCACCTTCTC CCCATAG    1137

```

(75) INFORMATION FOR SEQ ID NO:74:

```

25  (i) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 378 amino acids
      (B) TYPE: amino acid
      (C) STRANDEDNESS:
      (D) TOPOLOGY: not relevant

```

(ii) MOLECULE TYPE: protein

```

30  (xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

```

Met Asp Leu Gly Lys Pro Met Lys Ser Val Leu Val Val Ala Leu Leu
1 5 10 15

Val Ile Phe Gln Val Cys Leu Cys Cln Asp Glu Val Thr Asp Asp Tyr
20 25 30

5 Ile Gly Asp Asn Thr Thr Val Asp Tyr Thr Leu Phe Glu Ser Leu Cys
35 40 45

Ser Lys Lys Asp Val Arg Asn Phe Lys Ala Trp Phe Leu Pro Ile Met
50 55 60

10 Tyr Ser Ile Ile Cys Phe Val Gly Leu Leu Gly Asn Gly Leu Val Val
65 70 75 80

Leu Thr Tyr Ile Tyr Phe Lys Arg Leu Lys Thr Met Thr Asp Thr Tyr
85 90 95

Leu Leu Asn Leu Ala Val Ala Asp Ile Leu Phe Leu Leu Thr Leu Pro
100 105 110

15 Phe Trp Ala Tyr Ser Ala Ala Lys Ser Trp Val Phe Gly Val His Phe
115 120 125

Cys Lys Leu Ile Phe Ala Ile Tyr Lys Met Ser Phe Phe Ser Gly Met
130 135 140

20 Leu Leu Leu Leu Cys Ile Ser Ile Asp Arg Tyr Val Ala Ile Val Gln
145 150 155 160

Ala Val Ser Ala His Arg His Arg Ala Arg Val Leu Leu Ile Ser Lys
165 170 175

Leu Ser Cys Val Gly Ile Trp Ile Leu Ala Thr Val Leu Ser Ile Pro
180 185 190

25 Glu Leu Leu Tyr Ser Asp Leu Gln Arg Ser Ser Ser Glu Gln Ala Met
195 200 205

Arg Cys Ser Leu Ile Thr Glu His Val Glu Ala Phe Ile Thr Ile Gln
210 215 220

30 Val Ala Gln Met Val Ile Gly Phe Leu Val Pro Leu Leu Ala Met Ser
225 230 235 240

Phe Cys Tyr Leu Val Ile Ile Arg Thr Leu Leu Gln Ala Arg Asn Phe
245 250 255

Glu Arg Asn Lys Ala Ile Lys Val Ile Ile Ala Val Val Val Val Phe
260 265 270

35 Ile Val Phe Gln Leu Pro Tyr Asn Gly Val Val Leu Ala Gln Thr Val
275 280 285

59

Ala Asn Phe Asn Ile Thr Ser Ser Thr Cys Glu Leu Ser Lys Gln Leu
290 295 300

Asn Ile Ala Tyr Asp Val Thr Tyr Ser Leu Ala Cys Val Arg Cys Cys
305 310 315 320

5 Val Asn Pro Phe Leu Tyr Ala Phe Ile Gly Val Lys Phe Arg Asn Asp
325 330 335

Leu Phe Lys Leu Phe Lys Asp Leu Gly Cys Leu Ser Gln Glu Gln Leu
340 345 350

10 Arg Gln Trp Ser Ser Cys Arg His Ile Arg Arg Ser Ser Met Ser Val
355 360 365

Glu Ala Glu Thr Thr Thr Thr Phe Ser Pro
370 375

(76) INFORMATION FOR SEQ ID NO:75:

15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

CTGGAATTCA CCTGGACCAC CACCAATGGA TA 32

(77) INFORMATION FOR SEQ ID NO:76:

25 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

30 CTCGGATCCT GCAAAGTTTG TCATACAGTT 30

(78) INFORMATION FOR SEQ ID NO:77:

35 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1085 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

	ATGGATATAC	AAATGGCAAA	CAATTTTACT	CCGCCCTCTG	CAACTCCTCA	GGGAAATGAC	60
	TGTGACCTCT	ATGCACATCA	CAGCACGGCC	AGGATAGTAA	TGCCTCTGCA	TTACAGCCTC	120
	GTCTTCATCA	TTGGGCTCGT	GGGAAACTTA	CTAGCCTTGG	TCGTCAATTGT	TCAAAACAGG	180
5	AAAAAAATCA	ACTCTACCAC	CCTCTATTCA	ACAAATTTGG	TGATTTCTGA	TATACTTTTT	240
	ACCACGGCTT	TGCCTACACG	AATAGCCTAC	TATGCAATGG	GCTTTGACTG	GAGAATCGGA	300
	GATGCCTTGT	GTAGGATAAC	TGCGCTAGTG	TTTTACATCA	ACACATATGC	AGGTGTGAAC	360
	TTTATGACCT	GCCTGAGTAT	TGACCGCTTC	ATTGCTGTGG	TGCACCCTCT	ACGCTACAAC	420
	AAGATAAAAA	GGATTGAACA	TGCAAAAGGC	GTGTGCATAT	TTGTCTGGAT	TCTAGTATTT	480
10	GCTCAGACAC	TCCCACTCCT	CATCAACCCT	ATGTCAAAGC	AGGAGGCTGA	AAGGATTACA	540
	TGCATGGAGT	ATCCAAACTT	TGAAGAACT	AAATCTCTTC	CCTGGATTCT	GCTTGGGGCA	600
	TGTTTTCATAG	GATATGTACT	TCCACTTATA	ATCATTCTCA	TCTGCTATTG	TCAGATCTGC	660
	TGCAAACTCT	TCAGAACTGC	CAAACAAAAC	CCACTCACTG	AGAAATCTGG	TGTAAACAAA	720
	AAGGCTCTCA	ACACAATTAT	TCTTATTATT	GTTGTGTTTG	TTCTCTGTTT	CACACCTTAC	780
15	CATGTTGCAA	TTATTCAACA	TATGATTAAG	AAGCTTCGTT	TCTCTAATTT	CCTGGAATGT	840
	AGCCAAAGAC	ATTCGTTCCA	GATTTCTCTG	CACTTTACAG	TATGCCTGAT	GAACCTCAAT	900
	TGCTGCATGG	ACCCTTTTAT	CTACTTCTTT	GCATGTAAAG	GGTATAAGAG	AAAGGTTATG	960
	AGGATGCTGA	AACGGCAAGT	CAGTGTATCG	ATTTCTAGTG	CTGTGAAGTC	AGCCCCTGAA	1020
	GAAAATTCAC	GTGAAATGAC	AGAAACGCAG	ATGATGATAC	ATTCCAAGTC	TTCAAATGGA	1080
20	AAGTGA						1086

(79) INFORMATION FOR SEQ ID NO:78:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 361 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

30 Met Asp Ile Gln Met Ala Asn Asn Phe Thr Pro Pro Ser Ala Thr Pro
1 5 10 15

61

Gln Gly Asn Asp Cys Asp Leu Tyr Ala His His Ser Thr Ala Arg Ile
 20 25 30
 Val Met Pro Leu His Tyr Ser Leu Val Phe Ile Ile Gly Leu Val Gly
 35 40 45
 5 Asn Leu Leu Ala Leu Val Val Ile Val Gln Asn Arg Lys Lys Ile Asn
 50 55 60
 Ser Thr Thr Leu Tyr Ser Thr Asn Leu Val Ile Ser Asp Ile Leu Phe
 65 70 75 80
 10 Thr Thr Ala Leu Pro Thr Arg Ile Ala Tyr Tyr Ala Met Gly Phe Asp
 85 90 95
 Trp Arg Ile Gly Asp Ala Leu Cys Arg Ile Thr Ala Leu Val Phe Tyr
 100 105 110
 Ile Asn Thr Tyr Ala Gly Val Asn Phe Met Thr Cys Leu Ser Ile Asp
 115 120 125
 15 Arg Phe Ile Ala Val Val His Pro Leu Arg Tyr Asn Lys Ile Lys Arg
 130 135 140
 Ile Glu His Ala Lys Gly Val Cys Ile Phe Val Trp Ile Leu Val Phe
 145 150 155 160
 20 Ala Gln Thr Leu Pro Leu Leu Ile Asn Pro Met Ser Lys Gln Glu Ala
 165 170 175
 Glu Arg Ile Thr Cys Met Glu Tyr Pro Asn Phe Glu Glu Thr Lys Ser
 180 185 190
 Leu Pro Trp Ile Leu Leu Gly Ala Cys Phe Ile Gly Tyr Val Leu Pro
 195 200 205
 25 Leu Ile Ile Ile Leu Ile Cys Tyr Ser Gln Ile Cys Cys Lys Leu Phe
 210 215 220
 Arg Thr Ala Lys Gln Asn Pro Leu Thr Glu Lys Ser Gly Val Asn Lys
 225 230 235 240
 30 Lys Ala Leu Asn Thr Ile Ile Leu Ile Ile Val Val Phe Val Leu Cys
 245 250 255
 Phe Thr Pro Tyr His Val Ala Ile Ile Gln His Met Ile Lys Lys Leu
 260 265 270
 Arg Phe Ser Asn Phe Leu Glu Cys Ser Gln Arg His Ser Phe Gln Ile
 275 280 285
 35 Ser Leu His Phe Thr Val Cys Leu Met Asn Phe Asn Cys Cys Met Asp
 290 295 300
 Pro Phe Ile Tyr Phe Phe Ala Cys Lys Gly Tyr Lys Arg Lys Val Met

5

10

- (ii) MOLECULE TYPE: DNA (genomic)

CTGGAATTCT CCTGCTCATC CAGCCATGCG G

31

20

- (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

25 CCTGGATCCC CACCCCTACT GGGGCCTCAG

30

30

- (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

ATGCGGTGGC TGTGGCCCT GGCTGTCTCT CTTGCTGTGA TTTTGGCTGT GGGGCTAAGC 60

35 AGGGTCTCTG GGGGTGCCCC CCTGCACCTG GGCAGGCACA GAGCCGAGAC CCAGGAGCAG 120

	CAGAGCCGAT CCAAGAGGGG CACCGAGGAT GAGGAGGCCA AGGGCGTGCA GCAGTATGTG	180
	CCTGAGGAGT GGGCGGAGTA CCCCCGGCCC ATTCACCCTG CTGGCCTGCA GCCAACCAAG	240
	CCCTTGGTGG CCACCAGCCC TAACCCCGAC AAGGATGGGG GCACCCGAGA CAGTGGGCAG	300
	GAACTGAGGG GCAATCTGAC AGGGGCACCA GGGCAGAGGC TACAGATCCA GAACCCCTG	360
5	TATCCGGTGA CCGAGAGCTC CTACAGTGCC TATGCCATCA TGCTTCTGGC GCTGGTGGTG	420
	TTTGCGGTGG GCATTGTGGG CAACCTGTGC GTCATGTGCA TCGTGTGGCA CAGCTACTAC	480
	CTGAAGAGCG CCTGGAAGTC CATCCTTGCC AGCCTGGCCC TCTGGGATTT TCTGGTCCTC	540
	TTTTTCTGCC TCCCTATTGT CATCTTCAAC GAGATCACCA AGCAGAGGCT ACTGGGTGAC	600
	GTTTCTTGTC GTGCCGTGCC CTTTCATGGAG GTCTCCTCTC TGGGAGTCAC GACTTTCAGC	660
10	CTCTGTGCCC TGGGCATTGA CCGCTTCCAC GTGGCCACCA GCACCCTGCC CAAGGTGAGG	720
	CCCATCGAGC GGTGCCAATC CATCCTGGCC AAGTTGGCTG TCATCTGGGT GGGCTCCATG	780
	ACGCTGGCTG TGCCTGAGCT CCTGCTGTGG CAGCTGGCAC AGGAGCCTGC CCCCACCATG	840
	GGCACCTGG ACTCATGCAT CATGAAACCC TCAGCCAGCC TGCCCGAGTC CCTGTATTCA	900
	CTGGTGATGA CCTACCAGAA CGCCCGCATG TGGTGGTACT TTGGCTGCTA CTTCTGCCTG	960
15	CCCATCCTCT TCACAGTCAC CTGCCAGCTG GTGACATGGC GGGTGCGAGG CCCTCCAGGG	1020
	AGGAAGTCAG AGTGCAGGGC CAGCAAGCAC GAGCAGTGTG AGAGCCAGCT CAACAGCACC	1080
	GTGGTGGGCC TGACCGTGGT CTACGCCTTC TGCACCCTCC CAGAGAACGT CTGCAACATC	1140
	GTGGTGGCCT ACCTCTCCAC CGAGCTGACC CGCCAGACCC TGGACCTCCT GGGCCTCATC	1200
	AACCAGTTCT CCACCTTCTT CAAGGGCGCC ATCACCCAG TGCTGCTCCT TTGCATCTGC	1260
20	AGGCCGCTGG GCCAGGCCTT CCTGGACTGC TGCTGCTGCT GCTGCTGTGA GGAGTGCGGC	1320
	GGGGCTTCGG AGGCCTCTGC TGCCAATGGG TCGGACAACA AGCTCAAGAC CGAGGTGTCC	1380
	TCTTCCATCT ACTTCCACAA GCCCAGGGAG TCACCCCCAC TCCTGCCCCT GGGCACACCT	1440
	TGCTGA	1446

(83) INFORMATION FOR SEQ ID NO:82:

- 25 (i). SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 481 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
- 30 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

Met Arg Trp Leu Trp Pro Leu Ala Val Ser Leu Ala Val Ile Leu Ala
 1 5 10 15
 Val Gly Leu Ser Arg Val Ser Gly Gly Ala Pro Leu His Leu Gly Arg
 5 20 25 30
 His Arg Ala Glu Thr Gln Glu Gln Gln Ser Arg Ser Lys Arg Gly Thr
 35 40 45
 Glu Asp Glu Glu Ala Lys Gly Val Gln Gln Tyr Val Pro Glu Glu Trp
 50 55 60
 Ala Glu Tyr Pro Arg Pro Ile His Pro Ala Gly Leu Gln Pro Thr Lys
 10 65 70 75 80
 Pro Leu Val Ala Thr Ser Pro Asn Pro Asp Lys Asp Gly Gly Thr Pro
 85 90 95
 Asp Ser Gly Gln Glu Leu Arg Gly Asn Leu Thr Gly Ala Pro Gly Gln
 15 100 105 110
 Arg Leu Gln Ile Gln Asn Pro Leu Tyr Pro Val Thr Glu Ser Ser Tyr
 115 120 125
 Ser Ala Tyr Ala Ile Met Leu Leu Ala Leu Val Val Phe Ala Val Gly
 130 135 140
 Ile Val Gly Asn Leu Ser Val Met Cys Ile Val Trp His Ser Tyr Tyr
 20 145 150 155 160
 Leu Lys Ser Ala Trp Asn Ser Ile Leu Ala Ser Leu Ala Leu Trp Asp
 165 170 175
 Phe Leu Val Leu Phe Phe Cys Leu Pro Ile Val Ile Phe Asn Glu Ile
 25 180 185 190
 Thr Lys Gln Arg Leu Leu Gly Asp Val Ser Cys Arg Ala Val Pro Phe
 195 200 205
 Met Glu Val Ser Ser Leu Gly Val Thr Thr Phe Ser Leu Cys Ala Leu
 210 215 220
 Gly Ile Asp Arg Phe His Val Ala Thr Ser Thr Leu Pro Lys Val Arg
 30 225 230 235 240
 Pro Ile Glu Arg Cys Gln Ser Ile Leu Ala Lys Leu Ala Val Ile Trp
 245 250 255
 Val Gly Ser Met Thr Leu Ala Val Pro Glu Leu Leu Leu Trp Gln Leu
 35 260 265 270
 Ala Gln Glu Pro Ala Pro Thr Met Gly Thr Leu Asp Ser Cys Ile Met

65

	275	280	285
	Lys Pro Ser Ala Ser Leu	Pro Glu Ser Leu Tyr	Ser Leu Val Met Thr
	290	295	300
5	Tyr Gln Asn Ala Arg Met Trp Trp Tyr Phe Gly Cys Tyr Phe Cys Leu		
	305	310	315 320
	Pro Ile Leu Phe Thr Val Thr Cys Gln Leu Val Thr Trp Arg Val Arg		
		325 330	335
	Gly Pro Pro Gly Arg Lys Ser Glu Cys Arg Ala Ser Lys His Glu Gln		
		340 345	350
10	Cys Glu Ser Gln Leu Asn Ser Thr Val Val Gly Leu Thr Val Val Tyr		
		355 360	365
	Ala Phe Cys Thr Leu Pro Glu Asn Val Cys Asn Ile Val Val Ala Tyr		
		370 375	380
15	Leu Ser Thr Glu Leu Thr Arg Gln Thr Leu Asp Leu Leu Gly Leu Ile		
		385 390	395 400
	Asn Gln Phe Ser Thr Phe Phe Lys Gly Ala Ile Thr Pro Val Leu Leu		
		405 410	415
	Leu Cys Ile Cys Arg Pro Leu Gly Gln Ala Phe Leu Asp Cys Cys Cys		
		420 425	430
20	Cys Cys Cys Cys Glu Glu Cys Gly Gly Ala Ser Glu Ala Ser Ala Ala		
		435 440	445
	Asn Gly Ser Asp Asn Lys Leu Lys Thr Glu Val Ser Ser Ser Ile Tyr		
		450 455	460
25	Phe His Lys Pro Arg Glu Ser Pro Pro Leu Leu Pro Leu Gly Thr Pro		
		465 470 475	480
	Cys		

(84) INFORMATION FOR SEQ ID NO:83:

- 30 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- 35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

ATGTGGAACG CGACGCCAG CG

(85) INFORMATION FOR SEQ ID NO:84:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

TCATGTATTA ATACTAGATT CT

22

10 (86) INFORMATION FOR SEQ ID NO:85:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 38 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

TACCATGTGG AACGCGACGC CCAGCGAAGA GCCGGGGT

38

(87) INFORMATION FOR SEQ ID NO:86:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

CGGAATTCAT GTATTAATAC TAGATTCTGT CCAGGCCCG

39

(88) INFORMATION FOR SEQ ID NO:87:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1101 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

ATGTGGAACG CGACGCCCAG CGAAGAGCCG GGGTTCAACC TCACACTGGC CGACCTGGAC 60
 TGGGATGCTT CCCCCGGCAA CGACTCGCTG GGCACGAGC TGCTGCAGCT CTTCCCCGCG 120
 CCGCTGCTGG CGGGCGTCAC AGCCACCTGC GTGGCACTCT TCGTGGTGGG TATCGCTGGC 180
 AACCTGCTCA CCATGCTGGT GGTGTCGCGC TTCCGCGAGC TGCGCACCAC CACCAACCTC 240
 5 TACCTGTCCA GCATGGCCTT CTCCGATCTG CTCATCTTCC TCTGCATGCC CCTGGACCTC 300
 GTTCGCCTCT GGCAGTACCG GCCCTGGAAC TTCGGCGACC TCCTCTGCAA ACTCTTCCAA 360
 TTCGTCAGTG AGAGCTGCAC CTACGCCACG GTGCTCACCA TCACAGCGCT GAGCGTCGAG 420
 CGCTACTTCG CCATCTGCTT CCCACTCCGG GCCAAGGTGG TGGTCACCAA GGGGCGGGTG 480
 AAGCTGGTCA TCTTCGTCAT CTGGGCGGTG GCCTTCTGCA GCGCCGGGCC CATCTTCGTG 540
 10 CTAGTCGGGG TGGAGCACGA GAACGGCACC GACCCTTGGG ACACCAACGA GTGCCGCCCC 600
 ACCGAGTTTG CGGTGCGCTC TGGACTGCTC ACGGTCATGG TGTGGGTGTC CAGCATCTTC 660
 TTCTTCCTTC CTGTCTTCTG TCTCACGGTC CTCTACAGTC TCATCGGCAG GAAGCTGTGG 720
 CGGAGGAGGC GCGGCGATGC TGTCGTGGGT GCCTCGCTCA GGGACCAGAA CCACAAGCAA 780
 ACCGTGAAAA TGCTGGCTGT AGTGGTGTTT GCCTTCATCC TCTGCTGGCT CCCCTTCCAC 840
 15 GTAGGGCGAT ATTTATTTTC CAAATCCTTT GAGCCTGGCT CTTGGAGAT TGCTCAGATC 900
 AGCCAGTACT GCAACCTCGT GTCCTTTGTC CTCTTCTACC TCAGTGCTGC CATCAACCCC 960
 ATTCTGTACA ACATCATGTC CAAGAAGTAC CGGGTGGCAG TGTTCAGACT TCTGGGATTC 1020
 GAACCCTTCT CCCAGAGAAA GCTCTCCACT CTGAAAGATG AAAGTTCTCG GGCCTGGACA 1080
 GAATCTAGTA TTAATACATG A 1101

20 (89) INFORMATION FOR SEQ ID NO:88:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 366 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

25 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

Met Trp Asn Ala Thr Pro Ser Glu Glu Pro Gly Phe Asn Leu Thr Leu
 1 5 10 15

30 Ala Asp Leu Asp Trp Asp Ala Ser Pro Gly Asn Asp Ser Leu Gly Asp
 20 25 30

Glu Leu Leu Gln Leu Phe Pro Ala Pro Leu Leu Ala Gly Val Thr Ala
 35 40 45
 Thr Cys Val Ala Leu Phe Val Val Gly Ile Ala Gly Asn Leu Leu Thr
 50 55 60
 5 Met Leu Val Val Ser Arg Phe Arg Glu Leu Arg Thr Thr Thr Asn Leu
 65 70 75 80
 Tyr Leu Ser Ser Met Ala Phe Ser Asp Leu Leu Ile Phe Leu Cys Met
 85 90 95
 10 Pro Leu Asp Leu Val Arg Leu Trp Gln Tyr Arg Pro Trp Asn Phe Gly
 100 105 110
 Asp Leu Leu Cys Lys Leu Phe Gln Phe Val Ser Glu Ser Cys Thr Tyr
 115 120 125
 Ala Thr Val Leu Thr Ile Thr Ala Leu Ser Val Glu Arg Tyr Phe Ala
 130 135 140
 15 Ile Cys Phe Pro Leu Arg Ala Lys Val Val Val Thr Lys Gly Arg Val
 145 150 155 160
 Lys Leu Val Ile Phe Val Ile Trp Ala Val Ala Phe Cys Ser Ala Gly
 165 170 175
 20 Pro Ile Phe Val Leu Val Gly Val Glu His Glu Asn Gly Thr Asp Pro
 180 185 190
 Trp Asp Thr Asn Glu Cys Arg Pro Thr Glu Phe Ala Val Arg Ser Gly
 195 200 205
 Leu Leu Thr Val Met Val Trp Val Ser Ser Ile Phe Phe Phe Leu Pro
 210 215 220
 25 Val Phe Cys Leu Thr Val Leu Tyr Ser Leu Ile Gly Arg Lys Leu Trp
 225 230 235 240
 Arg Arg Arg Arg Gly Asp Ala Val Val Gly Ala Ser Leu Arg Asp Gln
 245 250 255
 30 Asn His Lys Gln Thr Val Lys Met Leu Ala Val Val Val Phe Ala Phe
 260 265 270
 Ile Leu Cys Trp Leu Pro Phe His Val Gly Arg Tyr Leu Phe Ser Lys
 275 280 285
 Ser Phe Glu Pro Gly Ser Leu Glu Ile Ala Gln Ile Ser Gln Tyr Cys
 290 295 300
 35 Asn Leu Val Ser Phe Val Leu Phe Tyr Leu Ser Ala Ala Ile Asn Pro
 305 310 315 320
 Ile Leu Tyr Asn Ile Met Ser Lys Lys Tyr Arg Val Ala Val Phe Arg

Asp Glu Ser Ser Arg Ala Trp Thr Glu Ser Ser Ile Asn Thr
355 360 365

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

33

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

30

25 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1842 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:

35 AATTCTGCAA GAGACGTTCT GCGAGCCCGA GCACCCAGGG AGGAGCAGGG GGCAGCGTTT 240

CTTGCGGGAC CCTCCTGGGA CCTGCCGGCG GCCCCGGGCC GTGACCCGGC TGCAGGCAGA 300
 GGGGCGGAGG CGTCGGCAGC CGGACCCCCG GGACCTCCAA CCAGGCCACC TGGCCCCTGG 360
 AGGTGGAAG GTGCTCGGGG TCAGGAGCCT TCTGAACTT TGGGGAGAGG GAACCCACG 420
 GCCCTCCAGC TCTTCCTTCA GATCTCAGAG GAGGAAGAGA AGGGTCCCAG AGGCGCTGGC 480
 5 ATTTCCGGGC GTAGCCAGGA GCAGAGTGTG AAGACAGTCC CCGGAGCCAG CGATCTTTTT 540
 TACTGGCCAA GGAGAGCCGG GAAACTCCAG GGTTCACACC ACAAGCCCCT GTCCAAGACG 600
 GCCAATGGAC TGGCGGGGCA CGAAGGGTGG ACAATTGCAC TCCCGGGCCG GGCCTGGCC 660
 CAGAATGGAT CCTTGGGTGA AGGAATCCAT GAGCCTGGGG GTCCCCGCCG GGGAAACAGC 720
 ACGAACCGGC GTGTGAGACT GAAGAACCCC TTCTACCCGC TGACCCAGGA GTCCTATGGA 780
 10 GCCTACGCGG TCATGTGTCT GTCCGTGGTG ATCTTCGGGA CCGGCATCAT TGGCAACCTG 840
 GCGGTGATGA GCATCGTGTG CCACAACCTAC TACATGCGGA GCATCTCCAA CTCCCTCTTG 900
 GCCAACCTGG CCTTCTGGGA CTTTCTCATC ATCTTCTTCT GCCTTCGCT GGTCTCTTC 960
 CACGAGCTGA CCAAGAAGTG GCTGCTGGAG GACTTCTCCT GCAAGATCGT GCCCTATATA 1020
 GAGGTGCTT CTCTGGGAGT CACCACTTTC ACCTTATGTG CTCTGTGCAT AGACCGCTTC 1080
 15 CGTGCTGCCA CCAACGTACA GATGTACTAC GAAATGATCG AAACTGTTC CTCAACAAC 1140
 GCCAACTTG CTGTTATATG GGTGGGAGCT CTATTGTTAG CACTTCAGA AGTTGTTCTC 1200
 CGCCAGCTGA GCAAGGAGGA TTTGGGGTTT AGTGGCCGAG CTCCGGCAGA AAGGTGCATT 1260
 ATTAAGATCT CTCCTGATTT ACCAGACACC ATCTATGTTC TAGCCCTCAC CTACGACAGT 1320
 GCGAGACTGT GGTGGTATTT TGGCTGTTAC TTTTGTTCG CACGCTTTT CACCATCACC 1380
 20 TGCTCTCTAG TGA CTGCGAG GAAAATCCGC AAAGCAGAGA AAGCCTGTAC CCGAGGGAAT 1440
 AAACGGCAGA TTCAACTAGA GAGTCAGATG AACTGTACAG TAGTGGCACT GACCATTTTA 1500
 TATGGATTTT GCATTATTCC TGAAAATATC TGCAACATTG TTACTGCCTA CATGGCTACA 1560
 GGGGTTTCAC AGCAGACAAT GGACCTCCTT AATATCATCA GCCAGTTCCT TTTGTTCTTT 1620
 AAGTCCTGTG TCACCCAGT CCTCCTTTTC TGTCTTGCA AACCTTCAG TCGGGCCTTC 1680
 25 ATGGAGTGCT GCTGCTGTTG CTGTGAGGAA TGCATTCAGA AGTCTTCAAC GGTGACCAGT 1740
 GATGACAATG ACAACGAGTA CACCACGGAA CTCGAACTCT CGCCTTTCAG TACCATACGC 1800
 CGTGAAATGT CCACTTTTGC TTCTGTCGGA ACTCATTGCT GA 1842

(93) INFORMATION FOR SEQ ID NO:92:

71

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 613 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:92:

5 Met Arg Ala Pro Gly Ala Leu Leu Ala Arg Met Ser Arg Leu Leu Leu
 1 5 10 15
 10 Leu Leu Leu Leu Lys Val Ser Ala Ser Ser Ala Leu Gly Val Ala Pro
 20 25 30
 Ala Ser Arg Asn Glu Thr Cys Leu Gly Glu Ser Cys Ala Pro Thr Val
 35 40 45
 15 Ile Gln Arg Arg Gly Arg Asp Ala Trp Gly Pro Gly Asn Ser Ala Arg
 50 55 60
 Asp Val Leu Arg Ala Arg Ala Pro Arg Glu Glu Gln Gly Ala Ala Phe
 65 70 75 80
 Leu Ala Gly Pro Ser Trp Asp Leu Pro Ala Ala Pro Gly Arg Asp Pro
 85 90 95
 20 Ala Ala Gly Arg Gly Ala Glu Ala Ser Ala Ala Gly Pro Pro Gly Pro
 100 105 110
 Pro Thr Arg Pro Pro Gly Pro Trp Arg Trp Lys Gly Ala Arg Gly Gln
 115 120 125
 25 Glu Pro Ser Glu Thr Leu Gly Arg Gly Asn Pro Thr Ala Leu Gln Leu
 130 135 140
 Phe Leu Gln Ile Ser Glu Glu Glu Glu Lys Gly Pro Arg Gly Ala Gly
 145 150 155 160
 Ile Ser Gly Arg Ser Gln Glu Gln Ser Val Lys Thr Val Pro Gly Ala
 165 170 175
 30 Ser Asp Leu Phe Tyr Trp Pro Arg Arg Ala Gly Lys Leu Gln Gly Ser
 180 185 190
 His His Lys Pro Leu Ser Lys Thr Ala Asn Gly Leu Ala Gly His Glu
 195 200 205
 35 Gly Trp Thr Ile Ala Leu Pro Gly Arg Ala Leu Ala Gln Asn Gly Ser
 210 215 220
 Leu Gly Glu Gly Ile His Glu Pro Gly Gly Pro Arg Arg Gly Asn Ser
 225 230 235 240

Thr Asn Arg Arg Val Arg Leu Lys Asn Pro Phe Tyr Pro Leu Thr Gln
 245 250 255
 Glu Ser Tyr Gly Ala Tyr Ala Val Met Cys Leu Ser Val Val Ile Phe
 260 265 270
 5 Gly Thr Gly Ile Ile Gly Asn Leu Ala Val Met Ser Ile Val Cys His
 275 280 285
 Asn Tyr Tyr Met Arg Ser Ile Ser Asn Ser Leu Leu Ala Asn Leu Ala
 290 295 300
 10 Phe Trp Asp Phe Leu Ile Ile Phe Phe Cys Leu Pro Leu Val Ile Phe
 305 310 315 320
 His Glu Leu Thr Lys Lys Trp Leu Leu Glu Asp Phe Ser Cys Lys Ile
 325 330 335
 Val Pro Tyr Ile Glu Val Ala Ser Leu Gly Val Thr Thr Phe Thr Leu
 340 345 350
 15 Cys Ala Leu Cys Ile Asp Arg Phe Arg Ala Ala Thr Asn Val Gln Met
 355 360 365
 Tyr Tyr Glu Met Ile Glu Asn Cys Ser Ser Thr Thr Ala Lys Leu Ala
 370 375 380
 20 Val Ile Trp Val Gly Ala Leu Leu Leu Ala Leu Pro Glu Val Val Leu
 385 390 395 400
 Arg Gln Leu Ser Lys Glu Asp Leu Gly Phe Ser Gly Arg Ala Pro Ala
 405 410 415
 Glu Arg Cys Ile Ile Lys Ile Ser Pro Asp Leu Pro Asp Thr Ile Tyr
 420 425 430
 25 Val Leu Ala Leu Thr Tyr Asp Ser Ala Arg Leu Trp Trp Tyr Phe Gly
 435 440 445
 Cys Tyr Phe Cys Leu Pro Thr Leu Phe Thr Ile Thr Cys Ser Leu Val
 450 455 460
 30 Thr Ala Arg Lys Ile Arg Lys Ala Glu Lys Ala Cys Thr Arg Gly Asn
 465 470 475 480
 Lys Arg Gln Ile Gln Leu Glu Ser Gln Met Asn Cys Thr Val Val Ala
 485 490 495
 Leu Thr Ile Leu Tyr Gly Phe Cys Ile Ile Pro Glu Asn Ile Cys Asn
 500 505 510
 35 Ile Val Thr Ala Tyr Met Ala Thr Gly Val Ser Gln Gln Thr Met Asp
 515 520 525
 Leu Leu Asn Ile Ile Ser Gln Phe Leu Leu Phe Phe Lys Ser Cys Val

73

530 535 540

Thr Pro Val Leu Leu Phe Cys Leu Cys Lys Pro Phe Ser Arg Ala Phe
545 550 555 560

5 Met Glu Cys Cys Cys Cys Cys Cys Glu Glu Cys Ile Gln Lys Ser Ser
565 570 575

Thr Val Thr Ser Asp Asp Asn Asp Asn Glu Tyr Thr Thr Glu Leu Glu
580 585 590

Leu Ser Pro Phe Ser Thr Ile Arg Arg Glu Met Ser Thr Phe Ala Ser
595 600 605

10 Val Gly Thr His Cys
610

(94) INFORMATION FOR SEQ ID NO:93:

- 15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:93:

20 CAGAATTCAG AGAAAAAAG TGAATATGGT TTTT 34

(95) INFORMATION FOR SEQ ID NO:94:

- 25 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:

TTGGATCCCT GGTGCATAAC AATTGAAAGA AT 32

30 (96) INFORMATION FOR SEQ ID NO:95:

- 35 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1248 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:

ATGGTTTTTG CTCACAGAAT GGATAACAGC AAGCCACATT TGATTATTCC TACACTTCTG 60
 GTGCCCCCTCC AAAACCGCAG CTGCACTGAA ACAGCCACAC CTCTGCCAAG CCAATACCTG 120
 ATGGAATTAA GTGAGGAGCA CAGTTGGATG AGCAACCAA CAGACCTTCA CTATGTGCTG 180
 5 AAACCCGGGG AAGTGGCCAC AGCCAGCATC TTCTTTGGGA TTCTGTGGTT GTTTTCTATC 240
 TTCGGCAATT CCCTGGTTTG TTTGGTCATC CATAGGAGTA GGAGGACTCA GTCTACCACC 300
 AACTACTTTG TGGTCTCCAT GGCATGTGCT GACCTTCTCA TCAGCGTTGC CAGCACGCCT 360
 TTCGTCTGCT TCCAGTTCAC CACTGGAAGG TGGACGCTGG GTAGTGCAAC GTGCAAGGTT 420
 GTGCGATATT TTCAATATCT CACTCCAGGT GTCCAGATCT ACGTTCTCCT CTCCATCTGC 480
 10 ATAGACCGGT TCTACACCAT CGTCTATCCT CTGAGCTTCA AGGTGTCCAG AGAAAAAGCC 540
 AAGAAAATGA TTGCGGCATC GTGGATCTTT GATGCAGGCT TTGTGACCCC TGTGCTCTTT 600
 TTCTATGGCT CCAACTGGGA CAGTCATTGT AACTATTTCC TCCCCTCCTC TTGGGAAGGC 660
 ACTGCCTACA CTGTCATCCA CTTCTTGGTG GGCTTTGTGA TTCCATCTGT CCTCATAATT 720
 TTATTTTACC AAAAGGTCAT AAAATATATT TGGAGAATAG GCACAGATGG CCGAACGGTG 780
 15 AGGAGGACAA TGAACATTGT CCCTCGGACA AAAGTGAAAA CTATCAAGAT GTTCCTCATT 840
 TTAAATCTGT TGTTTTTGCT CTCCTGGCTG CCTTTTCATG TAGCTCAGCT ATGGCACCCC 900
 CATGAACAAG ACTATAAGAA AAGTCCCTT GTTTTCACAG CTATCACATG GATATCCTTT 960
 AGTTCTTCAG CCTCTAAACC TACTCTGTAT TCAATTTATA ATGCCAATTT TCGGAGAGGG 1020
 ATGAAAGAGA CTTTTTGCAT GTCCTCTATG AAATGTTACC GAAGCAATGC CTATACTATC 1080
 20 ACAACAAGTT CAAGGATGGC CAAAAAAAC TACGTTGGCA TTTCAGAAAT CCCTTCCATG 1140
 GCCAAACTA TTACCAAAGA CTCGATCTAT GACTCATTTG ACAGAGAAGC CAAGGAAAAA 1200
 AAGCTTGCTT GGCCCATTA CTCAATCCA CCAAATACTT TTGTCTAA 1248

(97) INFORMATION FOR SEQ ID NO:96:

- 25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 415 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

- 30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:

75

Met Val Phe Ala His Arg Met Asp Asn Ser Lys Pro His Leu Ile Ile
 1 5 10 15
 Pro Thr Leu Leu Val Pro Leu Gln Asn Arg Ser Cys Thr Glu Thr Ala
 20 25 30
 5 Thr Pro Leu Pro Ser Gln Tyr Leu Met Glu Leu Ser Glu Glu His Ser
 35 40 45
 Trp Met Ser Asn Gln Thr Asp Leu His Tyr Val Leu Lys Pro Gly Glu
 50 55 60
 10 Val Ala Thr Ala Ser Ile Phe Phe Gly Ile Leu Trp Leu Phe Ser Ile
 65 70 75 80
 Phe Gly Asn Ser Leu Val Cys Leu Val Ile His Arg Ser Arg Arg Thr
 85 90 95
 Gln Ser Thr Thr Asn Tyr Phe Val Val Ser Met Ala Cys Ala Asp Leu
 100 105 110
 15 Leu Ile Ser Val Ala Ser Thr Pro Phe Val Leu Leu Gln Phe Thr Thr
 115 120 125
 Gly Arg Trp Thr Leu Gly Ser Ala Thr Cys Lys Val Val Arg Tyr Phe
 130 135 140
 20 Gln Tyr Leu Thr Pro Gly Val Gln Ile Tyr Val Leu Leu Ser Ile Cys
 145 150 155 160
 Ile Asp Arg Phe Tyr Thr Ile Val Tyr Pro Leu Ser Phe Lys Val Ser
 165 170 175
 Arg Glu Lys Ala Lys Lys Met Ile Ala Ala Ser Trp Ile Phe Asp Ala
 180 185 190
 25 Gly Phe Val Thr Pro Val Leu Phe Phe Tyr Gly Ser Asn Trp Asp Ser
 195 200 205
 His Cys Asn Tyr Phe Leu Pro Ser Ser Trp Glu Gly Thr Ala Tyr Thr
 210 215 220
 30 Val Ile His Phe Leu Val Gly Phe Val Ile Pro Ser Val Leu Ile Ile
 225 230 235 240
 Leu Phe Tyr Gln Lys Val Ile Lys Tyr Ile Trp Arg Ile Gly Thr Asp
 245 250 255
 Gly Arg Thr Val Arg Arg Thr Met Asn Ile Val Pro Arg Thr Lys Val
 260 265 270
 35 Lys Thr Ile Lys Met Phe Leu Ile Leu Asn Leu Leu Phe Leu Leu Ser
 275 280 285

Trp Leu Pro Phe His Val Ala Gln Leu Trp His Pro His Glu Gln Asp
 290 295 300

Tyr Lys Lys Ser Ser Leu Val Phe Thr Ala Ile Thr Trp Ile Ser Phe
 305 310 315 320

5 Ser Ser Ser Ala Ser Lys Pro Thr Leu Tyr Ser Ile Tyr Asn Ala Asn
 325 330 335

Phe Arg Arg Gly Met Lys Glu Thr Phe Cys Met Ser Ser Met Lys Cys
 340 345 350

10 Tyr Arg Ser Asn Ala Tyr Thr Ile Thr Thr Ser Ser Arg Met Ala Lys
 355 360 365

Lys Asn Tyr Val Gly Ile Ser Glu Ile Pro Ser Met Ala Lys Thr Ile
 370 375 380

Thr Lys Asp Ser Ile Tyr Asp Ser Phe Asp Arg Glu Ala Lys Glu Lys
 385 390 395 400

15 Lys Leu Ala Trp Pro Ile Asn Ser Asn Pro Pro Asn Thr Phe Val
 405 410 415

(98) INFORMATION FOR SEQ ID NO:97:

20 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:97:

25 GGAAAGCTTA ACGATCCCCA GGAGCAACAT

30

(99) INFORMATION FOR SEQ ID NO:98:

30 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 31 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:98:

CTGGGATCCT ACGAGAGCAT TTTTCACACA G

31

35 (100) INFORMATION FOR SEQ ID NO:99:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1842 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:99:

	ATGGGGCCCA CCCTAGCGGT TCCCACCCCC TATGGCTGTA TTGGCTGTAA GCTACCCCAG	60
	CCAGAATACC CACCGGCTCT AATCATCTTT ATGTTCTGCG CGATGGTTAT CACCATCGTT	120
	GTAGACCTAA TCGGCAACTC CATGGTCATT TTGGCTGTGA CGAAGAACAA GAAGCTCCGG	180
10	AATTCTGGCA ACATCTTCGT GGTCACTCTC TCTGTGGCCG ATATGCTGGT GGCCATCTAC	240
	CCATACCCTT TGATGCTGCA TGCCATGTCC ATTGGGGGCT GGGATCTGAG CCAGTTACAG	300
	TGCCAGATGG TCGGGTTCAT CACAGGGCTG AGTGTGGTCG GCTCCATCTT CAACATCGTG	360
	GCAATCGCTA TCAACCGTTA CTGCTACATC TGCCACAGCC TCCAGTACGA ACGGATCTTC	420
	AGTGTGCGCA ATACCTGCAT CTACCTGGTC ATCACCTGGA TCATGACCGT CCTGGCTGTC	480
15	CTGCCCAACA TGTACATTGG CACCATCGAG TACGATCCTC GCACCTACAC CTGCATCTTC	540
	AACTATCTGA ACAACCCTGT CTTCACTGTT ACCATCGTCT GCATCCACTT CGTCCTCCCT	600
	CTCCTCATCG TGGGTTTCTG CTACGTGAGG ATCTGGACCA AAGTGCTGGC GGCCCGTGAC	660
	CCTGCAGGGC AGAATCCTGA CAACCAACTT GCTGAGGTTT GCAATTTTCT AACCATGTTT	720
	GTGATCTTCC TCCTCTTTGC AGTGTGCTGG TGCCCTATCA ACGTGCTCAC TGTCTTGGTG	780
20	GCTGTCAGTC CGAAGGAGAT GGCAGGCAAG ATCCCCAACT GGCTTTATCT TGCAGCCTAC	840
	TTCATAGCCT ACTTCAACAG CTGCCTCAAC GCTGTGATCT ACGGGCTCCT CAATGAGAAT	900
	TTCCGAAGAG AATACTGGAC CATCTTCCAT GCTATGCGGC ACCCTATCAT ATTCTTCCCT	960
	GGCCTCATCA GTGATATTCG TGAGATGCAG GAGGCCCCGTA CCCTGGCCCCG CGCCCGTGCC	1020
	CATGCTCGCG ACCAAGCTCG TGAACAAGAC CGTGCCCATG CCTGTCCTGC TGTGGAGGAA	1080
25	ACCCCGATGA ATGTCCGGAA TGTTCATTA CCTGGTGATG CTGCAGCTGG CCACCCCGAC	1140
	CGTGCCTCTG GCCACCCTAA GCCCCATTCC AGATCCTCCT CTGCCTATCG CAAATCTGCC	1200
	TCTACCCACC ACAAGTCTGT CTTTAGCCAC TCCAAGGCTG CCTCTGGTCA CCTCAAGCCT	1260
	GTCTCTGGCC ACTCCAAGCC TGCCTCTGGT CACCCCAAGT CTGCCACTGT CTACCCTAAG	1320
	CCTGCCTCTG TCCATTTCAA GGGTGACTCT GTCCATTTCA AGGGTGACTC TGTCCATTTT	1380

AAGCCTGACT CTGTTTCATTT CAAGCCTGCT TCCAGCAACC CCAAGCCCAT CACTGGCCAC 1440
 CATGTCTCTG CTGGCAGCCA CTCCAAGTCT GCCTTCAGTG CTGCCACCAG CCACCCTAAA 1500
 CCCATCAAGC CAGCTACCAG CCATGCTGAG CCCACCACTG CTGACTATCC CAAGCCTGCC 1560
 ACTACCAGCC ACCCTAAGCC CGCTGCTGCT GACAACCCTG AGCTCTCTGC CTCCCATTGC 1620
 5 CCCGAGATCC CTGCCATTGC CCACCCTGTG TCTGACGACA GTGACCTCCC TGAGTCGGCC 1680
 TCTAGCCCTG CCGCTGGGCC CACCAAGCCT GCTGCCAGCC AGCTGGAGTC TGACACCATC 1740
 GCTGACCTTC CTGACCCTAC TGTAGTCACT ACCAGTACCA ATGATTACCA TGATGTCGTG 1800
 GTTGTGATG TTGAAGATGA TCCTGATGAA ATGGCTGTGT GA 1842

(101) INFORMATION FOR SEQ ID NO:100:

- 10 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 613 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant

- 15 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:

Met	Gly	Pro	Thr	Leu	Ala	Val	Pro	Thr	Pro	Tyr	Gly	Cys	Ile	Gly	Cys	1	5	10	15
Lys	Leu	Pro	Gln	Pro	Glu	Tyr	Pro	Pro	Ala	Leu	Ile	Ile	Phe	Met	Phe	20	25	30	
Cys	Ala	Met	Val	Ile	Thr	Ile	Val	Val	Asp	Leu	Ile	Gly	Asn	Ser	Met	35	40	45	
Val	Ile	Leu	Ala	Val	Thr	Lys	Asn	Lys	Lys	Leu	Arg	Asn	Ser	Gly	Asn	50	55	60	
Ile	Phe	Val	Val	Ser	Leu	Ser	Val	Ala	Asp	Met	Leu	Val	Ala	Ile	Tyr	65	70	75	80
Pro	Tyr	Pro	Leu	Met	Leu	His	Ala	Met	Ser	Ile	Gly	Gly	Trp	Asp	Leu	85	90	95	
Ser	Gln	Leu	Gln	Cys	Gln	Met	Val	Gly	Phe	Ile	Thr	Gly	Leu	Ser	Val	100	105	110	
Val	Gly	Ser	Ile	Phe	Asn	Ile	Val	Ala	Ile	Ala	Ile	Asn	Arg	Tyr	Cys	115	120	125	
Tyr	Ile	Cys	His	Ser	Leu	Gln	Tyr	Glu	Arg	Ile	Phe	Ser	Val	Arg	Asn	130	135	140	

79

	Thr	Cys	Ile	Tyr	Leu	Val	Ile	Thr	Trp	Ile	Met	Thr	Val	Leu	Ala	Val	
	145					150					155					160	
	Leu	Pro	Asn	Met	Tyr	Ile	Gly	Thr	Ile	Glu	Tyr	Asp	Pro	Arg	Thr	Tyr	
					165					170					175		
5	Thr	Cys	Ile	Phe	Asn	Tyr	Leu	Asn	Asn	Pro	Val	Phe	Thr	Val	Thr	Ile	
				180					185					190			
	Val	Cys	Ile	His	Phe	Val	Leu	Pro	Leu	Leu	Ile	Val	Gly	Phe	Cys	Tyr	
			195					200					205				
10	Val	Arg	Ile	Trp	Thr	Lys	Val	Leu	Ala	Ala	Arg	Asp	Pro	Ala	Gly	Gln	
		210					215					220					
	Asn	Pro	Asp	Asn	Gln	Leu	Ala	Glu	Val	Arg	Asn	Phe	Leu	Thr	Met	Phe	
		225				230					235					240	
	Val	Ile	Phe	Leu	Leu	Phe	Ala	Val	Cys	Trp	Cys	Pro	Ile	Asn	Val	Leu	
				245						250					255		
15	Thr	Val	Leu	Val	Ala	Val	Ser	Pro	Lys	Glu	Met	Ala	Gly	Lys	Ile	Pro	
			260						265					270			
	Asn	Trp	Leu	Tyr	Leu	Ala	Ala	Tyr	Phe	Ile	Ala	Tyr	Phe	Asn	Ser	Cys	
		275						280					285				
20	Leu	Asn	Ala	Val	Ile	Tyr	Gly	Leu	Leu	Asn	Glu	Asn	Phe	Arg	Arg	Glu	
		290					295					300					
	Tyr	Trp	Thr	Ile	Phe	His	Ala	Met	Arg	His	Pro	Ile	Ile	Phe	Phe	Pro	
		305				310					315					320	
	Gly	Leu	Ile	Ser	Asp	Ile	Arg	Glu	Met	Gln	Glu	Ala	Arg	Thr	Leu	Ala	
				325						330					335		
25	Arg	Ala	Arg	Ala	His	Ala	Arg	Asp	Gln	Ala	Arg	Glu	Gln	Asp	Arg	Ala	
			340						345					350			
	His	Ala	Cys	Pro	Ala	Val	Glu	Glu	Thr	Pro	Met	Asn	Val	Arg	Asn	Val	
			355					360					365				
30	Pro	Leu	Pro	Gly	Asp	Ala	Ala	Ala	Gly	His	Pro	Asp	Arg	Ala	Ser	Gly	
		370				375						380					
	His	Pro	Lys	Pro	His	Ser	Arg	Ser	Ser	Ser	Ala	Tyr	Arg	Lys	Ser	Ala	
		385				390					395					400	
	Ser	Thr	His	His	Lys	Ser	Val	Phe	Ser	His	Ser	Lys	Ala	Ala	Ser	Gly	
				405					410						415		
35	His	Leu	Lys	Pro	Val	Ser	Gly	His	Ser	Lys	Pro	Ala	Ser	Gly	His	Pro	
			420					425						430			
	Lys	Ser	Ala	Thr	Val	Tyr	Pro	Lys	Pro	Ala	Ser	Val	His	Phe	Lys	Gly	

	435	440	445
	Asp Ser Val His Phe Lys Gly Asp Ser Val His Phe Lys Pro Asp Ser 450 455 460		
5	Val His Phe Lys Pro Ala Ser Ser Asn Pro Lys Pro Ile Thr Gly His 465 470 475 480		
	His Val Ser Ala Gly Ser His Ser Lys Ser Ala Phe Ser Ala Ala Thr 485 490 495		
	Ser His Pro Lys Pro Ile Lys Pro Ala Thr Ser His Ala Glu Pro Thr 500 505 510		
10	Thr Ala Asp Tyr Pro Lys Pro Ala Thr Thr Ser His Pro Lys Pro Ala 515 520 525		
	Ala Ala Asp Asn Pro Glu Leu Ser Ala Ser His Cys Pro Glu Ile Pro 530 535 540		
15	Ala Ile Ala His Pro Val Ser Asp Asp Ser Asp Leu Pro Glu Ser Ala 545 550 555 560		
	Ser Ser Pro Ala Ala Gly Pro Thr Lys Pro Ala Ala Ser Gln Leu Glu 565 570 575		
	Ser Asp Thr Ile Ala Asp Leu Pro Asp Pro Thr Val Val Thr Thr Ser 580 585 590		
20	Thr Asn Asp Tyr His Asp Val Val Val Val Asp Val Glu Asp Asp Pro 595 600 605		
	Asp Glu Met Ala Val 610		

(102) INFORMATION FOR SEQ ID NO:101:

- 25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 32 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- 30 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:101:

TCCAAGCTTC GCCATGGGAC ATAACGGGAG CT

32

(103) INFORMATION FOR SEQ ID NO:102:

- 35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:102:

CGTGAATTCC AAGAATTAC AATCCTTGCT

30

5 (104) INFORMATION FOR SEQ ID NO:103:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1548 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:103:

	ATGGGACATA	ACGGGAGCTG	GATCTCTCCA	AATGCCAGCG	AGCCGCACAA	CGCGTCCGGC	60
	GCCGAGGCTG	CGGGTGTGAA	CCGCAGCGCG	CTCGGGGAGT	TCGGCGAGGC	GCAGCTGTAC	120
15	CGCCAGTTCA	CCACCACCGT	GCAGGTCGTC	ATCTTCATAG	GCTCGCTGCT	CGGAAACTTC	180
	ATGGTGTTAT	GGTCAACTTG	CCGCACAACC	GTGTTCAAAT	CTGTCACCAA	CAGGTTCAATT	240
	AAAAACCTGG	CCTGCTCGGG	GATTTGTGCC	AGCCTGGTCT	GTGTGCCCTT	CGACATCATC	300
	CTCAGCACCA	GTCCTCACTG	TTGCTGGTGG	ATCTACACCA	TGCTCTTCTG	CAAGGTCGTC	360
	AAATTTTTC	ACAAAGTATT	CTGCTCTGTG	ACCATCCTCA	GCTTCCCTGC	TATTGCTTTG	420
20	GACAGGTACT	ACTCAGTCCT	CTATCCACTG	GAGAGGAAAA	TATCTGATGC	CAAGTCCCGT	480
	GAACTGGTGA	TGTACATCTG	GGCCCATGCA	GTGGTGGCCA	GTGTCCCTGT	GTTTGCAGTA	540
	ACCAATGTGG	CTGACATCTA	TGCCACGTCC	ACCTGCACGG	AAGTCTGGAG	CAACTCCTTG	600
	GGCCACCTGG	TGTACGTTCT	GGTGTATAAC	ATCACCACGG	TCATTGTGCC	TGTGGTGGTG	660
	GTGTTCTCT	TCTTGATACT	GATCCGACGG	GCCCTGAGTG	CCAGCCAGAA	GAAGAAGGTC	720
25	ATCATAGCAG	CGCTCCGGAC	CCCACAGAAC	ACCATCTCTA	TTCCCTATGC	CTCCCAGCGG	780
	GAGGCCGAGC	TGCACGCCAC	CCTGCTCTCC	ATGGTGATGG	TCTTCATCTT	GTGTAGCGTG	840
	CCCTATGCCA	CCCTGGTCGT	CTACCAGACT	GTGCTCAATG	TCCCTGACAC	TTCCGTCTTC	900
	TTGCTGCTCA	CTGCTGTTTG	GCTGCCCAAA	GTCTCCCTGC	TGGCAAACCC	TGTTCTCTTT	960
	CTTACTGTGA	ACAAATCTGT	CCGCAAGTGC	TTGATAGGGA	CCCTGGTGCA	ACTACACCAC	1020
30	CGGTACAGTC	GCCGTAATGT	GGTCAGTACA	GGGAGTGGCA	TGGCTGAGGC	CAGCCTGGAA	1080

CCCAGCATAC GCTCGGGTAG CCAGCTCCTG GAGATGTTCC ACATTGGGCA GCAGCAGATC 1140
 TTTAAGCCCA CAGAGGATGA GGAAGAGAGT GAGGCCAAGT ACATTGGCTC AGCTGACTTC 1200
 CAGGCCAAGG AGATATTTAG CACCTGCCTC GAGGGAGAGC AGGGGCCACA GTTTGCGCCC 1260
 TCTGCCCCAC CCCTGAGCAC AGTGGACTCT GTATCCCAGG TGGCACCGGC AGCCCCTGTG 1320
 5 GAACCTGAAA CATTCCCTGA TAAGTATTCC CTGCAGTTTG GCTTTGGGCC TTTTGAGTTG 1380
 CCTCCTCAGT GGCTCTCAGA GACCCGAAAC AGCAAGAAGC GGCTGCTTCC CCCCTTGGGC 1440
 AACACCCAG AAGAGCTGAT CCAGACAAAG GTGCCCAAGG TAGGCAGGGT GGAGCGGAAG 1500
 ATGAGCAGAA ACAATAAAGT GAGCATTTTT CCAAAGGTGG ATTCCTAG 1548

(105) INFORMATION FOR SEQ ID NO:104:

- 10 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 515 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

- 15 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:104:

Met Gly His Asn Gly Ser Trp Ile Ser Pro Asn Ala Ser Glu Pro His
 1 5 10 15
 Asn Ala Ser Gly Ala Glu Ala Ala Gly Val Asn Arg Ser Ala Leu Gly
 20 25 30
 Glu Phe Gly Glu Ala Gln Leu Tyr Arg Gln Phe Thr Thr Thr Val Gln
 35 40 45
 Val Val Ile Phe Ile Gly Ser Leu Leu Gly Asn Phe Met Val Leu Trp
 50 55 60
 25 Ser Thr Cys Arg Thr Thr Val Phe Lys Ser Val Thr Asn Arg Phe Ile
 65 70 75 80
 Lys Asn Leu Ala Cys Ser Gly Ile Cys Ala Ser Leu Val Cys Val Pro
 85 90 95
 30 Phe Asp Ile Ile Leu Ser Thr Ser Pro His Cys Cys Trp Trp Ile Tyr
 100 105 110
 Thr Met Leu Phe Cys Lys Val Val Lys Phe Leu His Lys Val Phe Cys
 115 120 125
 Ser Val Thr Ile Leu Ser Phe Pro Ala Ile Ala Leu Asp Arg Tyr Tyr
 130 135 140

83

	Ser Val Leu Tyr Pro Leu Glu Arg Lys Ile Ser Asp Ala Lys Ser Arg	
	145	150 155 160
	Glu Leu Val Met Tyr Ile Trp Ala His Ala Val Val Ala Ser Val Pro	
		165 170 175
5	Val Phe Ala Val Thr Asn Val Ala Asp Ile Tyr Ala Thr Ser Thr Cys	
		180 185 190
	Thr Glu Val Trp Ser Asn Ser Leu Gly His Leu Val Tyr Val Leu Val	
		195 200 205
10	Tyr Asn Ile Thr Thr Val Ile Val Pro Val Val Val Val Phe Leu Phe	
		210 215 220
	Leu Ile Leu Ile Arg Arg Ala Leu Ser Ala Ser Gln Lys Lys Lys Val	
		225 230 235 240
	Ile Ile Ala Ala Leu Arg Thr Pro Gln Asn Thr Ile Ser Ile Pro Tyr	
		245 250 255
15	Ala Ser Gln Arg Glu Ala Glu Leu His Ala Thr Leu Leu Ser Met Val	
		260 265 270
	Met Val Phe Ile Leu Cys Ser Val Pro Tyr Ala Thr Leu Val Val Tyr	
		275 280 285
20	Gln Thr Val Leu Asn Val Pro Asp Thr Ser Val Phe Leu Leu Leu Thr	
		290 295 300
	Ala Val Trp Leu Pro Lys Val Ser Leu Leu Ala Asn Pro Val Leu Phe	
		305 310 315 320
	Leu Thr Val Asn Lys Ser Val Arg Lys Cys Leu Ile Gly Thr Leu Val	
		325 330 335
25	Gln Leu His His Arg Tyr Ser Arg Arg Asn Val Val Ser Thr Gly Ser	
		340 345 350
	Gly Met Ala Glu Ala Ser Leu Glu Pro Ser Ile Arg Ser Gly Ser Gln	
		355 360 365
30	Leu Leu Glu Met Phe His Ile Gly Gln Gln Gln Ile Phe Lys Pro Thr	
		370 375 380
	Glu Asp Glu Glu Glu Ser Glu Ala Lys Tyr Ile Gly Ser Ala Asp Phe	
		385 390 395 400
	Gln Ala Lys Glu Ile Phe Ser Thr Cys Leu Glu Gly Glu Gln Gly Pro	
		405 410 415
35	Gln Phe Ala Pro Ser Ala Pro Pro Leu Ser Thr Val Asp Ser Val Ser	
		420 425 430
	Gln Val Ala Pro Ala Ala Pro Val Glu Pro Glu Thr Phe Pro Asp Lys	

435 440 445

Tyr Ser Leu Gln Phe Gly Phe Gly Pro Phe Glu Leu Pro Pro Gln Trp
 450 455 460

5 Leu Ser Glu Thr Arg Asn Ser Lys Lys Arg Leu Leu Pro Pro Leu Gly
 465 470 475 480

Asn Thr Pro Glu Glu Leu Ile Gln Thr Lys Val Pro Lys Val Gly Arg
 485 490 495

Val Glu Arg Lys Met Ser Arg Asn Asn Lys Val Ser Ile Phe Pro Lys
 500 505 510

10 Val Asp Ser
 515

(106) INFORMATION FOR SEQ ID NO:105:

15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 29 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:105:

20 GGAGAATTCA CTAGGCGAGG CGCTCCATC

29

(107) INFORMATION FOR SEQ ID NO:106:

25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:106:

GGAGGATCCA GGAAACCTTA GGCCGAGTCC

30

30 (108) INFORMATION FOR SEQ ID NO:107:

35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1164 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:107:

ATGAATCGGC ACCATCTGCA GGATCACTTT CTGGAAATAG ACAAGAAGAA CTGCTGTGTG 60
TTCCGAGATG ACTTCATTGC CAAGGTGTTG CCGCCGGTGT TGGGGCTGGA GTTTATCTTT 120
GGGCTTCTGG GCAATGGCCT TGCCCTGTGG ATTTTCTGTT TCCACCTCAA GTCCTGGAAA 180
5 TCCAGCCGGA TTTTCCTGTT CAACCTGGCA GTAGCTGACT TTCTACTGAT CATCTGCCTG 240
CCGTTTCGTGA TGGACTACTA TGTGCGGCGT TCAGACTGGA ACTTTGGGGA CATCCCTTGC 300
CGGCTGGTGC TCTTCATGTT TGCCATGAAC CGCCAGGGCA GCATCATCTT CCTCACGGTG 360
GTGGCGGTAG ACAGGTATTT CCGGGTGGTC CATCCCCACC ACGCCCTGAA CAAGATCTCC 420
AATTGGACAG CAGCCATCAT CTCTTGCCTT CTGTGGGGCA TCACTGTTGG CCTAACAGTC 480
10 CACCTCCTGA AGAAGAAGTT GCTGATCCAG AATGGCCCTG CAAATGTGTG CATCAGCTTC 540
AGCATCTGCC ATACCTTCCG GTGGCAGGAA GCTATGTTCC TCCTGGAGTT CCTCCTGCCC 600
CTGGGCATCA TCCTGTTCTG CTCAGCCAGA ATTATCTGGA GCCTGCGGCA GAGACAAATG 660
GACCGGCATG CCAAGATCAA GAGAGCCATC ACCTTCATCA TGGTGGTGGC CATCGTCTTT 720
GTCATCTGCT TCCTTCCCAG CGTGGTTGTG CGGATCCGCA TCTTCTGGCT CCTGCACACT 780
15 TCGGGCACGC AGAATTGTGA AGTGTACCGC TCGGTGGACC TGGCGTTCTT TATCACTCTC 840
AGCTTCACCT ACATGAACAG CATGCTGGAC CCCGTGGTGT ACTACTTCTC CAGCCCATCC 900
TTTCCCAACT TCTTCTCCAC TTTGATCAAC CGCTGCCTCC AGAGGAAGAT GACAGGTGAG 960
CCAGATAATA ACCGCAGCAC GAGCGTCGAG CTCACAGGGG ACCCCAACAA AACCAGAGGC 1020
GCTCCAGAGG CGTTAATGGC CAACTCCGGT GAGCCATGGA GCCCCTCTTA TCTGGGCCCA 1080
20 ACCTCAAATA ACCATTCCAA GAAGGGACAT TGTCACCAAG AACCAGCATC TCTGGAGAAA 1140
CAGTTGGGCT GTTGCATCGA GTAA 1164

(109) INFORMATION FOR SEQ ID NO:108:

(i) SEQUENCE CHARACTERISTICS:

- 25 (A) LENGTH: 387 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:108:

30 Met Asn Arg His His Leu Gln Asp His Phe Leu Glu Ile Asp Lys Lys

1 5 10 15
 Asn Cys Cys Val Phe Arg Asp Asp Phe Ile Ala Lys Val Leu Pro Pro
 20 25 30
 5 Val Leu Gly Leu Glu Phe Ile Phe Gly Leu Leu Gly Asn Gly Leu Ala
 35 40 45
 Leu Trp Ile Phe Cys Phe His Leu Lys Ser Trp Lys Ser Ser Arg Ile
 50 55 60
 Phe Leu Phe Asn Leu Ala Val Ala Asp Phe Leu Leu Ile Ile Cys Leu
 65 70 75 80
 10 Pro Phe Val Met Asp Tyr Tyr Val Arg Arg Ser Asp Trp Asn Phe Gly
 85 90 95
 Asp Ile Pro Cys Arg Leu Val Leu Phe Met Phe Ala Met Asn Arg Gln
 100 105 110
 15 Gly Ser Ile Ile Phe Leu Thr Val Val Ala Val Asp Arg Tyr Phe Arg
 115 120 125
 Val Val His Pro His His Ala Leu Asn Lys Ile Ser Asn Trp Thr Ala
 130 135 140
 Ala Ile Ile Ser Cys Leu Leu Trp Gly Ile Thr Val Gly Leu Thr Val
 145 150 155 160
 20 His Leu Leu Lys Lys Lys Leu Leu Ile Gln Asn Gly Pro Ala Asn Val
 165 170 175
 Cys Ile Ser Phe Ser Ile Cys His Thr Phe Arg Trp His Glu Ala Met
 180 185 190
 25 Phe Leu Leu Glu Phe Leu Leu Pro Leu Gly Ile Ile Leu Phe Cys Ser
 195 200 205
 Ala Arg Ile Ile Trp Ser Leu Arg Gln Arg Gln Met Asp Arg His Ala
 210 215 220
 Lys Ile Lys Arg Ala Ile Thr Phe Ile Met Val Val Ala Ile Val Phe
 225 230 235 240
 30 Val Ile Cys Phe Leu Pro Ser Val Val Val Arg Ile Arg Ile Phe Trp
 245 250 255
 Leu Leu His Thr Ser Gly Thr Gln Asn Cys Glu Val Tyr Arg Ser Val
 260 265 270
 35 Asp Leu Ala Phe Phe Ile Thr Leu Ser Phe Thr Tyr Met Asn Ser Met
 275 280 285
 Leu Asp Pro Val Val Tyr Tyr Phe Ser Ser Pro Ser Phe Pro Asn Phe
 290 295 300

87

Phe Ser Thr Leu Ile Asn Arg Cys Leu Gln Arg Lys Met Thr Gly Glu
 305 310 315 320
 Pro Asp Asn Asn Arg Ser Thr Ser Val Glu Leu Thr Gly Asp Pro Asn
 325 330 335
 5 Lys Thr Arg Gly Ala Pro Glu Ala Leu Met Ala Asn Ser Gly Glu Pro
 340 345 350
 Trp Ser Pro Ser Tyr Leu Gly Pro Thr Ser Asn Asn His Ser Lys Lys
 355 360 365
 10 Gly His Cys His Gln Glu Pro Ala Ser Leu Glu Lys Gln Leu Gly Cys
 370 375 380
 Cys Ile Glu
 385

(110) INFORMATION FOR SEQ ID NO:109:

- 15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 37 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: DNA (genomic)
 20 (iv) ANTI-SENSE: NO
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:109:

ACCATGGCTT GCAATGGCAG TCGGCCAGG GGGCACT

37

(111) INFORMATION FOR SEQ ID NO:110:

- 25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 39 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: DNA (genomic)
 30 (iv) ANTI-SENSE: YES
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:110:

CGACCAGGAC AAACAGCATC TTGGTCACTT GTCTCCGGC

39

(112) INFORMATION FOR SEQ ID NO:111:

- 35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 39 base pairs
 (B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

5 (iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:111:

GACCAAGATG CTGTTTGTCC TGGTCGTGGT GTTTGGCAT

39

(113) INFORMATION FOR SEQ ID NO:112:

10 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

15 (iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:112:

CGGAATTCAG GATGGATCGG TCTCTTGCTG CGCCT

35

(114) INFORMATION FOR SEQ ID NO:113:

20 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1212 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:113:

ATGGCTTGCA ATGGCAGTGC GGCCAGGGGG CACTTTGACC CTGAGGACTT GAACCTGACT 60

GACGAGGCAC TGAGACTCAA GTACCTGGGG CCCAGCAGA CAGAGCTGTT CATGCCCATC 120

TGTGCCACAT ACCTGCTGAT CTTCGTGGTG GGCGCTGTGG GCAATGGGCT GACCTGTCTG 180

GTCATCCTGC GCCACAAGGC CATGCGCAGC CCTACCAACT ACTACCTCTT CAGCCTGGCC 240

30 GTGTCGGACC TGCTGGTGCT GCTGGTGGGC CTGCCCCTGG AGCTCTATGA GATGTGGCAC 300

AACTACCCCT TCCTGCTGGG CGTTGGTGGC TGCTATTTCC GCACGCTACT GTTTGAGATG 360

GTCTGCCTGG CCTCAGTGCT CAACGTCCT GCCCTGAGCG TGGAACGCTA TGTGGCCGTG 420

GTGCACCCAC TCCAGGCCAG GTCCATGGTG ACGCGGGCCC ATGTGCGCCG AGTGCTTGGG 480

GCCGTCTGGG GTCTTGCCAT GCTCTGCTCC CTGCCCCAACA CCAGCCTGCA CGGCATCCGG 540
 CAGCTGCACG TGCCCTGCCG GGGCCCAGTG CCAGACTCAG CTGTTTGCACT GCTGGTCCGC 600
 CCACGGGCCC TCTACAACAT GGTAGTGAG ACCACCGCGC TGCTCTTCTT CTGCCTGCCC 660
 ATGGCCATCA TGAGCGTGCT CTACCTGCTC ATTGGGCTGC GACTGCGGCG GGAGAGGCTG 720
 5 CTGCTCATGC AGGAGGCCAA GGGCAGGGGC TCTGCAGCAG CCAGGTCCAG ATACACCTGC 780
 AGGCTCCAGC AGCACGATCG GGGCCGGAGA CAAGTGACCA AGATGCTGTT TGTCTTGGTC 840
 GTGGTGTTTG GCATCTGCTG GGCCCCGTTT CACGCCGACC GCGTCATGTG GAGCGTCGTG 900
 TCACAGTGGA CAGATGGCCT GCACCTGGCC TTCCAGCACG TGCACGTCAT CTCCGGCATC 960
 TTCTTCTACC TGGGCTCGGC GGCCAACCCC GTGCTCTATA GCCTCATGTC CAGCCGCTTC 1020
 10 CGAGAGACCT TCCAGGAGGC CCTGTGCCTC GGGGCCTGCT GCCATCGCCT CAGACCCCGC 1080
 CACAGCTCCC ACAGCCTCAG CAGGATGACC ACAGGCAGCA CCCTGTGTGA TGTGGGCTCC 1140
 CTGGGCAGCT GGGTCCACCC CCTGGCTGGG AACGATGGCC CAGAGGCGCA GCAAGAGACC 1200
 GATCCATCCT GA 1212

(115) INFORMATION FOR SEQ ID NO:114:

- 15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 403 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

- 20 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:114:

Met Ala Cys Asn Gly Ser Ala Ala Arg Gly His Phe Asp Pro Glu Asp
 1 5 10 15
 Leu Asn Leu Thr Asp Glu Ala Leu Arg Leu Lys Tyr Leu Gly Pro Gln
 25 20 25 30
 Gln Thr Glu Leu Phe Met Pro Ile Cys Ala Thr Tyr Leu Leu Ile Phe
 35 40 45
 Val Val Gly Ala Val Gly Asn Gly Leu Thr Cys Leu Val Ile Leu Arg
 50 55 60
 30 His Lys Ala Met Arg Thr Pro Thr Asn Tyr Tyr Leu Phe Ser Leu Ala
 65 70 75 80
 Val Ser Asp Leu Leu Val Leu Leu Val Gly Leu Pro Leu Glu Leu Tyr
 85 90 95

	Glu Met Trp His Asn Tyr Pro Phe Leu Leu Gly Val Gly Gly Cys Tyr	
	100	105 110
	Phe Arg Thr Leu Leu Phe Glu Met Val Cys Leu Ala Ser Val Leu Asn	
	115	120 125
5	Val Thr Ala Leu Ser Val Glu Arg Tyr Val Ala Val Val His Pro Leu	
	130	135 140
	Gln Ala Arg Ser Met Val Thr Arg Ala His Val Arg Arg Val Leu Gly	
	145	150 155 160
10	Ala Val Trp Gly Leu Ala Met Leu Cys Ser Leu Pro Asn Thr Ser Leu	
	165	170 175
	His Gly Ile Arg Gln Leu His Val Pro Cys Arg Gly Pro Val Pro Asp	
	180	185 190
	Ser Ala Val Cys Met Leu Val Arg Pro Arg Ala Leu Tyr Asn Met Val	
	195	200 205
15	Val Gln Thr Thr Ala Leu Leu Phe Phe Cys Leu Pro Met Ala Ile Met	
	210	215 220
	Ser Val Leu Tyr Leu Leu Ile Gly Leu Arg Leu Arg Arg Glu Arg Leu	
	225	230 235 240
20	Leu Leu Met Gln Glu Ala Lys Gly Arg Gly Ser Ala Ala Ala Arg Ser	
	245	250 255
	Arg Tyr Thr Cys Arg Leu Gln Gln His Asp Arg Gly Arg Arg Gln Val	
	260	265 270
	Thr Lys Met Leu Phe Val Leu Val Val Val Phe Gly Ile Cys Trp Ala	
	275	280 285
25	Pro Phe His Ala Asp Arg Val Met Trp Ser Val Val Ser Gln Trp Thr	
	290	295 300
	Asp Gly Leu His Leu Ala Phe Gln His Val His Val Ile Ser Gly Ile	
	305	310 315 320
30	Phe Phe Tyr Leu Gly Ser Ala Ala Asn Pro Val Leu Tyr Ser Leu Met	
	325	330 335
	Ser Ser Arg Phe Arg Glu Thr Phe Gln Glu Ala Leu Cys Leu Gly Ala	
	340	345 350
	Cys Cys His Arg Leu Arg Pro Arg His Ser Ser His Ser Leu Ser Arg	
	355	360 365
35	Met Thr Thr Gly Ser Thr Leu Cys Asp Val Gly Ser Leu Gly Ser Trp	
	370	375 380
	Val His Pro Leu Ala Gly Asn Asp Gly Pro Glu Ala Gln Gln Glu Thr	

385

390

395

400

Asp Pro Ser

(116) INFORMATION FOR SEQ ID NO:115:

- 5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

- 10 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:115:

GGAAGCTTCA GGCCCAAAGA TGGGGAACAT

30

(117) INFORMATION FOR SEQ ID NO:116:

- 15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:116:

GTGGATCCAC CCGCGGAGGA CCCAGGCTAG

30

(118) INFORMATION FOR SEQ ID NO:117:

- 25 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1098 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:117:

30 ATGGGGAACA TCACTGCAGA CAACTCCTCG ATGAGCTGTA CCATCGACCA TACCATCCAC 60
CAGACGCTGG CCCCGGTGGT CTATGTTACC GTGCTGGTGG TGGGCTTCCC GGCCAACTGC 120
CTGTCCCTCT ACTTCGGCTA CCTGCAGATC AAGGCCCGGA ACGAGCTGGG CGTGTACCTG 180
TGCAACCTGA CGGTGGCCGA CCTCTTCTAC ATCTGCTCGC TGCCCTTCTG GCTGCAGTAC 240
GTGCTGCAGC ACGACAACCTG GTCTCACGGC GACCTGTCCT GCCAGGTGTG CGGCATCCTC 300
35 CTGTACGAGA ACATCTACAT CAGCGTGGGC TTCCTCTGCT GCATCTCCGT GGACCGCTAC 360

CTGGCTGTGG CCCATCCCTT CCGCTTCCAC CAGTTCCGGA CCCTGAAGGC GGCCGTCGGC 420
 GTCAGCGTGG TCATCTGGGC CAAGGAGCTG CTGACCAGCA TCTACTTCCT GATGCACGAG 480
 GAGGTCATCG AGGACGAGAA CCAGCACCGC GTGTGCTTTG AGCACTACCC CATCCAGGCA 540
 TGGCAGCGCG CCATCAACTA CTACCGCTTC CTGGTGGGCT TCCTCTTCCC CATCTGCCTG 600
 5 CTGCTGGCGT CCTACCAGGG CATCCTGCGC GCCGTGCGCC GGAGCCACGG CACCCAGAAG 660
 AGCCGCAAGG ACCAGATCCA GCGGCTGGTG CTCAGCACCG TGGTCATCTT CCTGGCCTGC 720
 TTCCTGCCCT ACCACGTGTT GCTGCTGGTG CGCAGCGTCT GGGAGGCCAG CTGCGACTTC 780
 GCCAAGGGCG TTTTCAACGC CTACCACTTC TCCCTCCTGC TCACCAGCTT CAACTGCGTC 840
 GCGGACCCCG TGCTCTACTG CTTCGTCAGC GAGACCACCC ACCGGGACCT GGCCCGCCTC 900
 10 CGCGGGGCCT GCCTGGCCTT CCTCACCTGC TCCAGGACCG GCCGGGCCAG GGAGGCCTAC 960
 CCGCTGGGTG CCCCCGAGGC CTCCGGGAAA AGCGGGGCCC AGGGTGAGGA GCCCGAGCTG 1020
 TTGACCAAGC TCCACCCGGC CTTCCAGACC CCTAACTCGC CAGGGTCGGG CGGGTTCCCC 1080
 ACGGGCAGGT TGGCCTAG 1098

(119) INFORMATION FOR SEQ ID NO:118:

15

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 365 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant

20

- (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:118:

Met Gly Asn Ile Thr Ala Asp Asn Ser Ser Met Ser Cys Thr Ile Asp
 1 5 10 15

25

His Thr Ile His Gln Thr Leu Ala Pro Val Val Tyr Val Thr Val Leu
 20 25 30

Val Val Gly Phe Pro Ala Asn Cys Leu Ser Leu Tyr Phe Gly Tyr Leu
 35 40 45

Gln Ile Lys Ala Arg Asn Glu Leu Gly Val Tyr Leu Cys Asn Leu Thr
 50 55 60

30

Val Ala Asp Leu Phe Tyr Ile Cys Ser Leu Pro Phe Trp Leu Gln Tyr
 65 70 75 80

Val Leu Gln His Asp Asn Trp Ser His Gly Asp Leu Ser Cys Gln Val
 85 90 95

93

Cys Gly Ile Leu Leu Tyr Glu Asn Ile Tyr Ile Ser Val Gly Phe Leu
 100 105 110
 Cys Cys Ile Ser Val Asp Arg Tyr Leu Ala Val Ala His Pro Phe Arg
 115 120 125
 5 Phe His Gln Phe Arg Thr Leu Lys Ala Ala Val Gly Val Ser Val Val
 130 135 140
 Ile Trp Ala Lys Glu Leu Leu Thr Ser Ile Tyr Phe Leu Met His Glu
 145 150 155 160
 10 Glu Val Ile Glu Asp Glu Asn Gln His Arg Val Cys Phe Glu His Tyr
 165 170 175
 Pro Ile Gln Ala Trp Gln Arg Ala Ile Asn Tyr Tyr Arg Phe Leu Val
 180 185 190
 Gly Phe Leu Phe Pro Ile Cys Leu Leu Leu Ala Ser Tyr Gln Gly Ile
 195 200 205
 15 Leu Arg Ala Val Arg Arg Ser His Gly Thr Gln Lys Ser Arg Lys Asp
 210 215 220
 Gln Ile Gln Arg Leu Val Leu Ser Thr Val Val Ile Phe Leu Ala Cys
 225 230 235 240
 20 Phe Leu Pro Tyr His Val Leu Leu Leu Val Arg Ser Val Trp Glu Ala
 245 250 255
 Ser Cys Asp Phe Ala Lys Gly Val Phe Asn Ala Tyr His Phe Ser Leu
 260 265 270
 Leu Leu Thr Ser Phe Asn Cys Val Ala Asp Pro Val Leu Tyr Cys Phe
 275 280 285
 25 Val Ser Glu Thr Thr His Arg Asp Leu Ala Arg Leu Arg Gly Ala Cys
 290 295 300
 Leu Ala Phe Leu Thr Cys Ser Arg Thr Gly Arg Ala Arg Glu Ala Tyr
 305 310 315 320
 30 Pro Leu Gly Ala Pro Glu Ala Ser Gly Lys Ser Gly Ala Gln Gly Glu
 325 330 335
 Glu Pro Glu Leu Leu Thr Lys Leu His Pro Ala Phe Gln Thr Pro Asn
 340 345 350
 Ser Pro Gly Ser Gly Gly Phe Pro Thr Gly Arg Leu Ala
 355 360 365

35 (120) INFORMATION FOR SEQ ID NO:119:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 base pairs

(B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:119:

GACCTCGAGT CTTTCTACAC CTCATC

26

(121) INFORMATION FOR SEQ ID NO:120:

(i) SEQUENCE CHARACTERISTICS:

10 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:120:

15 TGCTCTAGAT TCCAGATAGG TGAAACTTG

30

(122) INFORMATION FOR SEQ ID NO:121:

(i) SEQUENCE CHARACTERISTICS:

20 (A) LENGTH: 1416 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:121:

ATGGATATTC TTTGTGAAGA AAATACTTCT TTGAGCTCAA CTACGAACTC CCTAATGCAA 60

25 TTAAATGATG ACAACAGGCT CTACAGTAAT GACTTTAACT CCGGAGAAGC TAACACTTCT 120

GATGCATTTA ACTGGACAGT CGACTCTGAA AATCGAACCA ACCTTTCCTG TGAAGGGTGC 180

CTCTCACCGT CGTGTCTCTC CTTACTTCAT CTCCAGGAAA AAAACTGGTC TGCTTTACTG 240

ACAGCCGTAG TGATTATTCT AACTATTGCT GGAAACATAC TCGTCATCAT GGCAGTGTCC 300

CTAGAGAAAA AGCTGCAGAA TGCCACCAAC TATTTCTGTA TGTCATTGCT CATAGCTGAT 360

30 ATGCTGCTGG GTTTCCTTGT CATGCCCGTG TCCATGTTAA CCATCCTGTA TGGGTACCGG 420

TGGCCTCTGC CGAGCAAGCT TTGTGCAGTC TGGATTTACC TGGACGTGCT CTTCTCCACG 480

GCCTCCATCA TGCACCTCTG CGCCATCTCG CTGGACCGCT ACGTCGCCAT CCAGAATCCC 540

ATCCACCACA GCCGCTTCAA CTCCAGAACT AAGGCATTTT TGAAAATCAT TGCTGTTTGG 600

ACCATATCAG TAGGTATATC CATGCCAATA CCAGTCTTTG GGCTACAGGA CGATTCTGAAG 660
 GTCTTTAAGG AGGGGAGTTG CTTACTCGCC GATGATAACT TTGTCCTGAT CGGCTCTTTT 720
 GTGTCATTTT TCATTCCCTT AACCATCATG GTGATCACCT ACTTTCTAAC TATCAAGTCA 780
 CTCCAGAAAG AAGCTACTTT GTGTGTAAGT GATCTTGGCA CACGGGCCAA ATTAGCTTCT 840
 5 TTCAGCTTCC TCCCTCAGAG TTCTTTGTCT TCAGAAAAGC TCTTCCAGCG GTCGATCCAT 900
 AGGGAGCCAG GGTCCCTACAC AGGCAGGAGG ACTATGCAGT CCATCAGCAA TGAGCAAAAG 960
 GCATGCAAGG TGCTGGGCAT CGTCTTCTTC CTGTTTGTGG TGATGTGGTG CCCTTTCTTC 1020
 ATCACAAACA TCATGGCCGT CATCTGCAA GAGTCCTGCA ATGAGGATGT CATTGGGGCC 1080
 CTGCTCAATG TGTTTGTGTTG GATCGGTTAT CTCTCTTCAG CAGTCAACCC ACTAGTCTAC 1140
 10 ACACTGTTCA ACAAGACCTA TAGGTCAGCC TTTTCACGGT ATATTCAGTG TCAGTACAAG 1200
 GAAAACAAAA AACCATTGCA GTTAATTTTA GTGAACACAA TACCGGCTTT GGCCTACAAG 1260
 TCTAGCCAAC TTCAAATGGG ACAAAAAAAG AATTCAAAGC AAGATGCCAA GACAACAGAT 1320
 AATGACTGCT CAATGGTTGC TCTAGGAAAG CAGTATTCTG AAGAGGCTTC TAAAGACAAT 1380
 AGCGACGGAG TGAATGAAAA GGTGAGCTGT GTGTGA 1416

15 (123) INFORMATION FOR SEQ ID NO:122:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 471 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - 20 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:122:

	Met	Asp	Ile	Leu	Cys	Glu	Glu	Asn	Thr	Ser	Leu	Ser	Ser	Thr	Thr	Asn
	1				5					10					15	
25	Ser	Leu	Met	Gln	Leu	Asn	Asp	Asp	Asn	Arg	Leu	Tyr	Ser	Asn	Asp	Phe
				20					25					30		
	Asn	Ser	Gly	Glu	Ala	Asn	Thr	Ser	Asp	Ala	Phe	Asn	Trp	Thr	Val	Asp
			35				40					45				
30	Ser	Glu	Asn	Arg	Thr	Asn	Leu	Ser	Cys	Glu	Gly	Cys	Leu	Ser	Pro	Ser
	50					55						60				
	Cys	Leu	Ser	Leu	Leu	His	Leu	Gln	Glu	Lys	Asn	Trp	Ser	Ala	Leu	Leu
	65				70					75					80	

Thr Ala Val Val Ile Ile Leu Thr Ile Ala Gly Asn Ile Leu Val Ile
 85 90 95
 Met Ala Val Ser Leu Glu Lys Lys Leu Gln Asn Ala Thr Asn Tyr Phe
 100 105 110
 5 Leu Met Ser Leu Ala Ile Ala Asp Met Leu Leu Gly Phe Leu Val Met
 115 120 125
 Pro Val Ser Met Leu Thr Ile Leu Tyr Gly Tyr Arg Trp Pro Leu Pro
 130 135 140
 10 Ser Lys Leu Cys Ala Val Trp Ile Tyr Leu Asp Val Leu Phe Ser Thr
 145 150 155 160
 Ala Ser Ile Met His Leu Cys Ala Ile Ser Leu Asp Arg Tyr Val Ala
 165 170 175
 Ile Gln Asn Pro Ile His His Ser Arg Phe Asn Ser Arg Thr Lys Ala
 180 185 190
 15 Phe Leu Lys Ile Ile Ala Val Trp Thr Ile Ser Val Gly Ile Ser Met
 195 200 205
 Pro Ile Pro Val Phe Gly Leu Gln Asp Asp Ser Lys Val Phe Lys Glu
 210 215 220
 20 Gly Ser Cys Leu Leu Ala Asp Asp Asn Phe Val Leu Ile Gly Ser Phe
 225 230 235 240
 Val Ser Phe Phe Ile Pro Leu Thr Ile Met Val Ile Thr Tyr Phe Leu
 245 250 255
 Thr Ile Lys Ser Leu Gln Lys Glu Ala Thr Leu Cys Val Ser Asp Leu
 260 265 270
 25 Gly Thr Arg Ala Lys Leu Ala Ser Phe Ser Phe Leu Pro Gln Ser Ser
 275 280 285
 Leu Ser Ser Glu Lys Leu Phe Gln Arg Ser Ile His Arg Glu Pro Gly
 290 295 300
 30 Ser Tyr Thr Gly Arg Arg Thr Met Gln Ser Ile Ser Asn Glu Gln Lys
 305 310 315 320
 Ala Cys Lys Val Leu Gly Ile Val Phe Phe Leu Phe Val Val Met Trp
 325 330 335
 Cys Pro Phe Phe Ile Thr Asn Ile Met Ala Val Ile Cys Lys Glu Ser
 340 345 350
 35 Cys Asn Glu Asp Val Ile Gly Ala Leu Leu Asn Val Phe Val Trp Ile
 355 360 365
 Gly Tyr Leu Ser Ser Ala Val Asn Pro Leu Val Tyr Thr Leu Phe Asn

97

370 375 380
 Lys Thr Tyr Arg Ser Ala Phe Ser Arg Tyr Ile Gln Cys Gln Tyr Lys
 385 390 395 400
 5 Glu Asn Lys Lys Pro Leu Gln Leu Ile Leu Val Asn Thr Ile Pro Ala
 405 410 415
 Leu Ala Tyr Lys Ser Ser Gln Leu Gln Met Gly Gln Lys Lys Asn Ser
 420 425 430
 Lys Gln Asp Ala Lys Thr Thr Asp Asn Asp Cys Ser Met Val Ala Leu
 435 440 445
 10 Gly Lys Gln Tyr Ser Glu Glu Ala Ser Lys Asp Asn Ser Asp Gly Val
 450 455 460
 Asn Glu Lys Val Ser Cys Val
 465 470

(124) INFORMATION FOR SEQ ID NO:123:

- 15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 27 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- 20 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:123:

GACCTCGAGG TTGCTTAAGA CTGAAGC

27

(125) INFORMATION FOR SEQ ID NO:124:

- 25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 27 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:124:

ATTTCTAGAC ATATGTAGCT TGTACCG

27

(126) INFORMATION FOR SEQ ID NO:125:

- 35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1377 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:125:

	ATGGTGAACC TGAGGAATGC GGTGCATTCA TTCCTTGTGC ACCTAATTGG CCTATTGGTT	60
	TGGCAATGTG ATATTTCTGT GAGCCCAGTA GCAGCTATAG TAACTGACAT TTTCAATACC	120
5	TCCGATGGTG GACGCTTCAA ATTCCCAGAC GGGGTACAAA ACTGGCCAGC ACTTTCAATC	180
	GTCATCATAA TAATCATGAC AATAGGTGGC AACATCCTTG TGATCATGGC AGTAAGCATG	240
	GAAAAGAAAC TGCACAATGC CACCAATTAC TTCTTAATGT CCCTAGCCAT TGCTGATATG	300
	CTAGTGGGAC TACTTGTCAT GCCCCTGTCT CTCCTGGCAA TCCTTTATGA TTATGTCTGG	360
	CCACTACCTA GATATTTGTG CCCCCTCTGG ATTTCTTTAG ATGTTTATT TTCAACAGCG	420
10	TCCATCATGC ACCTCTGCGC TATATCGCTG GATCGGTATG TAGCAATACG TAATCCTATT	480
	GAGCATAGCC GTTTC AATTC GCGGACTAAG GCCATCATGA AGATTGCTAT TGTTTGGGCA	540
	ATTTCTATAG GTGTATCAGT TCCTATCCCT GTGATTGGAC TGAGGGACGA AGAAAAGGTG	600
	TTCGTGAACA ACACGACGTG CGTGCTCAAC GACCCAAATT TCGTTCTTAT TGGGTCCTTC	660
	GTAGCTTTCT TCATACCGCT GACGATTATG GTGATTACGT ATTGCCTGAC CATCTACGTT	720
15	CTGCGCCGAC AAGCTTTGAT GTTACTGCAC GGCCACACCG AGGAACCGCC TGGACTAAGT	780
	CTGGATTTCC TGAAGTGCTG CAAGAGGAAT ACGGCCGAGG AAGAGAACTC TGCAAACCTT	840
	AACCAAGACC AGAACGCACG CCGAAGAAAG AAGAAGGAGA GACGTCCTAG GGGCACCATG	900
	CAGGCTATCA ACAATGAAAG AAAAGCTTCG AAAGTCCTTG GGATTGTTTT CTTTGTGTTT	960
	CTGATCATGT GGTGCCCATT TTTCATTACC AATATTCTGT CTGTTCTTTG TGAGAAGTCC	1020
20	TGTAACCAAA AGCTCATGGA AAAGCTTCTG AATGTGTTTG TTTGGATTGG CTATGTTTGT	1080
	TCAGGAATCA ATCCTCTGGT GTATACTCTG TTCAACAAAA TTTACCGAAG GGCATTCTCC	1140
	AACTATTTGC GTTGCAATTA TAAGGTAGAG AAAAAGCCTC CTGTCAGGCA GATTCCAAGA	1200
	GTTGCCGCCA CTGCTTTGTC TGGGAGGGAG CTTAATGTTA ACATTTATCG GCATACCAAT	1260
	GAACCGGTGA TCGAGAAAGC CAGTGACAAT GAGCCCGGTA TAGAGATGCA AGTTGAGAAT	1320
25	TTAGAGTTAC CAGTAAATCC CTCCAGTGTG GTTAGCGAAA GGATTAGCAG TGTGTGA	1377

(127) INFORMATION FOR SEQ ID NO:126:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 458 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:126:

5	Met Val Asn Leu Arg Asn Ala Val His Ser Phe Leu Val His Leu Ile	1	5	10	15
	Gly Leu Leu Val Trp Gln Cys Asp Ile Ser Val Ser Pro Val Ala Ala	20	25	30	
10	Ile Val Thr Asp Ile Phe Asn Thr Ser Asp Gly Gly Arg Phe Lys Phe	35	40	45	
	Pro Asp Gly Val Gln Asn Trp Pro Ala Leu Ser Ile Val Ile Ile Ile	50	55	60	
	Ile Met Thr Ile Gly Gly Asn Ile Leu Val Ile Met Ala Val Ser Met	65	70	75	80
15	Glu Lys Lys Leu His Asn Ala Thr Asn Tyr Phe Leu Met Ser Leu Ala	85	90	95	
	Ile Ala Asp Met Leu Val Gly Leu Leu Val Met Pro Leu Ser Leu Leu	100	105	110	
20	Ala Ile Leu Tyr Asp Tyr Val Trp Pro Leu Pro Arg Tyr Leu Cys Pro	115	120	125	
	Val Trp Ile Ser Leu Asp Val Leu Phe Ser Thr Ala Ser Ile Met His	130	135	140	
	Leu Cys Ala Ile Ser Leu Asp Arg Tyr Val Ala Ile Arg Asn Pro Ile	145	150	155	160
25	Glu His Ser Arg Phe Asn Ser Arg Thr Lys Ala Ile Met Lys Ile Ala	165	170	175	
	Ile Val Trp Ala Ile Ser Ile Gly Val Ser Val Pro Ile Pro Val Ile	180	185	190	
30	Gly Leu Arg Asp Glu Glu Lys Val Phe Val Asn Asn Thr Thr Cys Val	195	200	205	
	Leu Asn Asp Pro Asn Phe Val Leu Ile Gly Ser Phe Val Ala Phe Phe	210	215	220	
	Ile Pro Leu Thr Ile Met Val Ile Thr Tyr Cys Leu Thr Ile Tyr Val	225	230	235	240
35	Leu Arg Arg Gln Ala Leu Met Leu Leu His Gly His Thr Glu Glu Pro	245	250	255	

Pro Gly Leu Ser Leu Asp Phe Leu Lys Cys Cys Lys Arg Asn Thr Ala
 260 265 270

Glu Glu Glu Asn Ser Ala Asn Pro Asn Gln Asp Gln Asn Ala Arg Arg
 275 280 285

5 Arg Lys Lys Lys Glu Arg Arg Pro Arg Gly Thr Met Gln Ala Ile Asn
 290 295 300

Asn Glu Arg Lys Ala Ser Lys Val Leu Gly Ile Val Phe Phe Val Phe
 305 310 315 320

10 Leu Ile Met Trp Cys Pro Phe Phe Ile Thr Asn Ile Leu Ser Val Leu
 325 330 335

Cys Glu Lys Ser Cys Asn Gln Lys Leu Met Glu Lys Leu Leu Asn Val
 340 345 350

Phe Val Trp Ile Gly Tyr Val Cys Ser Gly Ile Asn Pro Leu Val Tyr
 355 360 365

15 Thr Leu Phe Asn Lys Ile Tyr Arg Arg Ala Phe Ser Asn Tyr Leu Arg
 370 375 380

Cys Asn Tyr Lys Val Glu Lys Lys Pro Pro Val Arg Gln Ile Pro Arg
 385 390 395 400

20 Val Ala Ala Thr Ala Leu Ser Gly Arg Glu Leu Asn Val Asn Ile Tyr
 405 410 415

Arg His Thr Asn Glu Pro Val Ile Glu Lys Ala Ser Asp Asn Glu Pro
 420 425 430

Gly Ile Glu Met Gln Val Glu Asn Leu Glu Leu Pro Val Asn Pro Ser
 435 440 445

25 Ser Val Val Ser Glu Arg Ile Ser Ser Val
 450 455

(128) INFORMATION FOR SEQ ID NO:127:

- (i) SEQUENCE CHARACTERISTICS:
- 30 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:127:

35 GGTAAGCTTG GCAGTCCACG CCAGGCCTTC

30

(129) INFORMATION FOR SEQ ID NO:128:

101

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:128:

TCCGAATTCT CTGTAGACAC AAGGCTTTGG

30

(130) INFORMATION FOR SEQ ID NO:129:

10

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1068 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

15

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:129:

ATGGATCAGT TCCCTGAATC AGTGACAGAA AACTTTGAGT ACGATGATTT GGCTGAGGCC 60
TGTTATATTG GGGACATCGT GGTCTTTGGG ACTGTGTTCC TGTCCATATT CTACTCCGTC 120
ATCTTTGCCA TTGGCCTGGT GGGAAATTTG TTGGTAGTGT TTGCCCTCAC CAACAGCAAG 180
20 AAGCCCAAGA GTGTCACCGA CATTTACCTC CTGAACCTGG CTTGTCTGA TCTGCTGTTT 240
GTAGCCACTT TGCCCTTCTG GACTCACTAT TTGATAAATG AAAAGGGCCT CCACAATGCC 300
ATGTGCAAAT TCACTACCGC CTTCTTCTTC ATCGGCTTTT TTGGAAGCAT ATTCTTCATC 360
ACCGTCATCA GCATTGATAG GTACCTGGCC ATCGTCCTGG CCGCCAACTC CATGAACAAC 420
CGGACCGTGC AGCATGGCGT CACCATCAGC CTAGGCGTCT GGGCAGCAGC CATTTTGGTG 480
25 GCAGCACCCC AGTTCATGTT CACAAAGCAG AAAGAAAATG AATGCCTTGG TGA CTACCCC 540
GAGGTCCTCC AGGAAATCTG GCCCGTGCTC CGCAATGTGG AAACAAATTT TCTTGGCTTC 600
CTACTCCCCC TGCTCATTAT GAGTTATTGC TACTTCAGAA TCATCCAGAC GCTGTTTTCC 660
TGCAAGAACC ACAAGAAAGC CAAAGCCATT AACTGATCC TTCTGGTGGT CATCGTGTTT 720
TTCCTCTTCT GGACACCCTA CAACGTTATG ATTTTCCTGG AGACGCTTAA GCTCTATGAC 780
30 TTCTTTCCCA GTTGTGACAT GAGGAAGGAT CTGAGGCTGG CCCTCAGTGT GACTGAGACG 840
GTTGCATTTA GCCATTGTTG CCTGAATCCT CTCATCTATG CATTTGCTGG GGAGAAGTTC 900
AGAAGATACC TTTACCACCT GTATGGGAAA TGCCTGGCTG TCCTGTGTGG GCGCTCAGTC 960

CACGTTGATT TCTCCTCATC TGAATCACAA AGGAGCAGGC ATGGAAGTGT TCTGAGCAGC 1020

AATTTTACTT ACCACACGAG TGATGGAGAT GCATTGCTCC TTCTCTGA 1068

(131) INFORMATION FOR SEQ ID NO:130:

5 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 355 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:130:

Met	Asp	Gln	Phe	Pro	Glu	Ser	Val	Thr	Glu	Asn	Phe	Glu	Tyr	Asp	Asp	1	5	10	15
Leu	Ala	Glu	Ala	Cys	Tyr	Ile	Gly	Asp	Ile	Val	Val	Phe	Gly	Thr	Val	20	25	30	
Phe	Leu	Ser	Ile	Phe	Tyr	Ser	Val	Ile	Phe	Ala	Ile	Gly	Leu	Val	Gly	35	40	45	
Asn	Leu	Leu	Val	Val	Phe	Ala	Leu	Thr	Asn	Ser	Lys	Lys	Pro	Lys	Ser	50	55	60	
Val	Thr	Asp	Ile	Tyr	Leu	Leu	Asn	Leu	Ala	Leu	Ser	Asp	Leu	Leu	Phe	65	70	75	80
Val	Ala	Thr	Leu	Pro	Phe	Trp	Thr	His	Tyr	Leu	Ile	Asn	Glu	Lys	Gly	85	90	95	
Leu	His	Asn	Ala	Met	Cys	Lys	Phe	Thr	Thr	Ala	Phe	Phe	Phe	Ile	Gly	100	105	110	
Phe	Phe	Gly	Ser	Ile	Phe	Phe	Ile	Thr	Val	Ile	Ser	Ile	Asp	Arg	Tyr	115	120	125	
Leu	Ala	Ile	Val	Leu	Ala	Ala	Asn	Ser	Met	Asn	Asn	Arg	Thr	Val	Gln	130	135	140	
His	Gly	Val	Thr	Ile	Ser	Leu	Gly	Val	Trp	Ala	Ala	Ala	Ile	Leu	Val	145	150	155	160
Ala	Ala	Pro	Gln	Phe	Met	Phe	Thr	Lys	Gln	Lys	Glu	Asn	Glu	Cys	Leu	165	170	175	
Gly	Asp	Tyr	Pro	Glu	Val	Leu	Gln	Glu	Ile	Trp	Pro	Val	Leu	Arg	Asn	180	185	190	
Val	Glu	Thr	Asn	Phe	Leu	Gly	Phe	Leu	Leu	Pro	Leu	Leu	Ile	Met	Ser	195	200	205	

103

Tyr Cys Tyr Phe Arg Ile Ile Gln Thr Leu Phe Ser Cys Lys Asn His
 210 215 220
 Lys Lys Ala Lys Ala Ile Lys Leu Ile Leu Leu Val Val Ile Val Phe
 225 230 235 240
 5 Phe Leu Phe Trp Thr Pro Tyr Asn Val Met Ile Phe Leu Glu Thr Leu
 245 250 255
 Lys Leu Tyr Asp Phe Phe Pro Ser Cys Asp Met Arg Lys Asp Leu Arg
 260 265 270
 10 Leu Ala Leu Ser Val Thr Glu Thr Val Ala Phe Ser His Cys Cys Leu
 275 280 285
 Asn Pro Leu Ile Tyr Ala Phe Ala Gly Glu Lys Phe Arg Arg Tyr Leu
 290 295 300
 Tyr His Leu Tyr Gly Lys Cys Leu Ala Val Leu Cys Gly Arg Ser Val
 305 310 315 320
 15 His Val Asp Phe Ser Ser Ser Glu Ser Gln Arg Ser Arg His Gly Ser
 325 330 335
 Val Leu Ser Ser Asn Phe Thr Tyr His Thr Ser Asp Gly Asp Ala Leu
 340 345 350
 20 Leu Leu Leu
 355

(132) INFORMATION FOR SEQ ID NO:131:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 32 base pairs
 (B) TYPE: nucleic acid
 25 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:131:

GATCTCCAGT AGGCATAAGT GGACAATTCT GG

32

30 (133) INFORMATION FOR SEQ ID NO:132:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 35 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:132:

CTCCTTCGGT CTCCTATCG TTGTCAGAAG

30

(134) INFORMATION FOR SEQ ID NO:133:

- 5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:133:

10 AGAAGGCCAA GATCGCGCGG CTGGCCCTCA

30

(135) INFORMATION FOR SEQ ID NO:134:

- 15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:134:

CGGCGCCACC GCACGAAAAA GTCATCTTC

30

20 (136) INFORMATION FOR SEQ ID NO:135:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
25 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:135:

GCCAAGAAGC GGGTGAAGTT CCTGGTGGTG GCA

33

(137) INFORMATION FOR SEQ ID NO:136:

- 30 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- 35 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:136:

CAGGCGGAAG GTGAAAGTCC TGGTCCTCGT

30

(138) INFORMATION FOR SEQ ID NO:137:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:137:

CGGCGCCTGC GGGCCAAGCG GCTGGTGGTG GTG

33

(139) INFORMATION FOR SEQ ID NO:138:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:138:

CCAAGCACAA AGCCAAGAAA GTGACCATCA C

31

(140) INFORMATION FOR SEQ ID NO:139:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:139:

GCGCCGGCGC ACCAAATGCT TGCTGGTGGT

30

(141) INFORMATION FOR SEQ ID NO:140:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 41 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:140:

CAAAAAGCTG AAGAAATCTA AGAAGATCAT CTTTATTGTC G

41

(142) INFORMATION FOR SEQ ID NO:141:

5

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:141:

CAAGACCAAG GCAAAACGCA TGATCGCCAT

30

(143) INFORMATION FOR SEQ ID NO:142:

15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:142:

GTCAAGGAGA AGTCCAAAAG GATCATCATC

30

(144) INFORMATION FOR SEQ ID NO:143:

25

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:143:

30

CGCCGCGTGC GGGCCAAGCA GCTCCTGCTC

30

(145) INFORMATION FOR SEQ ID NO:144:

35

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 33 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

107

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:144:

CCTGATAAGC GCTATAAAAT GGTCTGTGTTT CGA

33

(146) INFORMATION FOR SEQ ID NO:145:

- 5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:145:

GAAAGACAAA AGAGAGTCAA GAGGATGTCT TTATTG

36

(147) INFORMATION FOR SEQ ID NO:146:

- 15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:146:

CGGAGAAAGA GGGTGAAACG CACAGCCATC GCC

33

(148) INFORMATION FOR SEQ ID NO:147:

- 25 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:147:

30 AAGCTTCAGC GGGCCAAGGC ACTGGTCACC

30

(149) INFORMATION FOR SEQ ID NO:148:

- 35 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:148:

CAGCGGCAGA AGGC^AAAAG GGTGGCCATC

30

(150) INFORMATION FOR SEQ ID NO:149:

- 5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:149:

CGGCAGAAGC CGAAGCGCAT GATCCTCGCG

30

(151) INFORMATION FOR SEQ ID NO:150:

- 15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:150:

GAGCGCAACA AGGCCAAAAA GGTGATCATC

30

(152) INFORMATION FOR SEQ ID NO:151:

- 25 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:151:

30 GGTGTAAACA AAAAGGCTAA AAACACAATT ATTCTTATT

39

(153) INFORMATION FOR SEQ ID NO:152:

- 35 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

109

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:152:

GAGAGCCAGC TCAAGAGCAC CGTGGTG

27

(154) INFORMATION FOR SEQ ID NO:153:

- 5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:153:

CCACAAGCAA ACCAAGAAAA TGCTGGCTGT

30

(155) INFORMATION FOR SEQ ID NO:154:

- 15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:154:

CATCAAGTGT ATCATGTGCC AAGTACGCCC

30

(156) INFORMATION FOR SEQ ID NO:155:

- 25 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:155:

30 CTAGAGAGTC AGATGAAGTG TACAGTAGTG GCAC

34

(157) INFORMATION FOR SEQ ID NO:156:

- 35 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:156:

CGGACAAAAG TGAAAACTAA AAAGATGTTC CTCATT

36

(158) INFORMATION FOR SEQ ID NO:157:

- 5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:157:

GCTGAGGTTC GCAATAAACT AACCATGTTT GTG

33

(159) INFORMATION FOR SEQ ID NO:158:

- 15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:158:

GGGAGGCCGA GCTGAAAGCC ACCCTGCTC

29

(160) INFORMATION FOR SEQ ID NO:159:

- 25 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:159:

30 CAAGATCAAG AGAGCCAAAA CCTTCATCAT G

31

(161) INFORMATION FOR SEQ ID NO:160:

- 35 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

111

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:160:

CCGGAGACAA GTGAAGAAGA TGCTGTTTGT C

31

(162) INFORMATION FOR SEQ ID NO:161:

- 5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:161:

GCAAGGACCA GATCAAGCGG CTGGTGCTCA

30

(163) INFORMATION FOR SEQ ID NO:162:

- 15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:162:

CAAGAAAGCC AAAGCCAAGA AACTGATCCT TCTG

34

(164) INFORMATION FOR SEQ ID NO:163:

- 25 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1068 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:163:

30 ATGGAAGATT TGGAGGAAAC ATTATTTGAA GAATTTGAAA ACTATTCCTA TGACCTAGAC 60
TATTACTCTC TGGAGTCTGA TTTGGAGGAG AAAGTCCAGC TGGGAGTTGT TCACTGGGTC 120
TCCCTGGTGT TATATTGTTT GGCTTTTGTG CTGGGAATTC CAGGAAATGC CATCGTCATT 180
TGGTTCACGG GGCTCAAGTG GAAGAAGACA GTCACCACTC TGTGGTTCCT CAATCTAGCC 240
ATTGCGGATT TCATTTTCT TCTCTTCTG CCCCTGTACA TCTCCTATGT GGCCATGAAT 300

```

TTCCACTGGC CCTTTGGCAT CTGGCTGTGC AAAGCCAATT CCTTCACTGC CCAGTTGAAC 360
ATGTTTGCCA GTGTTTTTTT CCTGACAGTG ATCAGCCTGG ACCACTATAT CCACTTGATC 420
CATCCTGTCT TATCTCATCG GCATCGAACC CTCAAGAACT CTCTGATTGT CATTATATTC 480
ATCTGGCTTT TGGCTTCTCT AATTGGCGGT CCTGCCCTGT ACTTCCGGGA CACTGTGGAG 540
5 TTCAATAATC ATACTCTTTG CTATAACAAT TTTCAGAAGC ATGATCCTGA CCTCACTTTG 600
ATCAGGCACC ATGTTCTGAC TTGGGTGAAA TTTATCATTG GCTATCTCTT CCCTTTGCTA 660
ACAATGAGTA TTTGCTACTT GTGTCTCATC TTCAAGGTGA AGAAGCGAAC AGTCCTGATC 720
TCCAGTAGGC ATAAGTGGAC AATTCTGGTT GTGGTTGTGG CCTTGTGGT TTGCTGGACT 780
CCTTATCACC TGTTTAGCAT TTGGGAGCTC ACCATTACACC ACAATAGCTA TTCCCACCAT 840
10 GTGATGCAGG CTGGAATCCC CCTCTCCACT GGTTTGGCAT TCCTCAATAG TTGCTTGAAC 900
CCCATCCTTT ATGTCTAAT TAGTAAGAAG TTCCAAGCTC GCTTCCGGTC CTCAGTTGCT 960
GAGATACTCA AGTACACACT GTGGGAAGTC AGCTGTTCTG GCACAGTGAG TGAACAGCTC 1020
AGGAACTCAG AAACCAAGAA TCTGTGTCTC CTGGAAACAG CTCAATAA 1068

```

(165) INFORMATION FOR SEQ ID NO:164:

```

15 (i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 355 amino acids
    (B) TYPE: amino acid
    (C) STRANDEDNESS:
    (D) TOPOLOGY: not relevant

20 (ii) MOLECULE TYPE: protein

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:164:

Met Glu Asp Leu Glu Glu Thr Leu Phe Glu Glu Phe Glu Asn Tyr Ser
1      5      10      15
Tyr Asp Leu Asp Tyr Tyr Ser Leu Glu Ser Asp Leu Glu Glu Lys Val
25      20      25      30
Gln Leu Gly Val Val His Trp Val Ser Leu Val Leu Tyr Cys Leu Ala
35      40      45
Phe Val Leu Gly Ile Pro Gly Asn Ala Ile Val Ile Trp Phe Thr Gly
50      55      60
30 Leu Lys Trp Lys Lys Thr Val Thr Thr Leu Trp Phe Leu Asn Leu Ala
65      70      75      80
Ile Ala Asp Phe Ile Phe Leu Leu Phe Leu Pro Leu Tyr Ile Ser Tyr
85      90      95

```

113

	Val	Ala	Met	Asn	Phe	His	Trp	Pro	Phe	Gly	Ile	Trp	Leu	Cys	Lys	Ala	
				100					105					110			
	Asn	Ser	Phe	Thr	Ala	Gln	Leu	Asn	Met	Phe	Ala	Ser	Val	Phe	Phe	Leu	
			115					120					125				
5	Thr	Val	Ile	Ser	Leu	Asp	His	Tyr	Ile	His	Leu	Ile	His	Pro	Val	Leu	
		130					135					140					
	Ser	His	Arg	His	Arg	Thr	Leu	Lys	Asn	Ser	Leu	Ile	Val	Ile	Ile	Phe	
	145					150					155					160	
10	Ile	Trp	Leu	Leu	Ala	Ser	Leu	Ile	Gly	Gly	Pro	Ala	Leu	Tyr	Phe	Arg	
					165					170					175		
	Asp	Thr	Val	Glu	Phe	Asn	Asn	His	Thr	Leu	Cys	Tyr	Asn	Asn	Phe	Gln	
				180					185					190			
	Lys	His	Asp	Pro	Asp	Leu	Thr	Leu	Ile	Arg	His	His	Val	Leu	Thr	Trp	
			195					200					205				
15	Val	Lys	Phe	Ile	Ile	Gly	Tyr	Leu	Phe	Pro	Leu	Leu	Thr	Met	Ser	Ile	
		210					215					220					
	Cys	Tyr	Leu	Cys	Leu	Ile	Phe	Lys	Val	Lys	Lys	Arg	Thr	Val	Leu	Ile	
	225					230					235					240	
20	Ser	Ser	Arg	His	Lys	Trp	Thr	Ile	Leu	Val	Val	Val	Val	Ala	Phe	Val	
					245					250					255		
	Val	Cys	Trp	Thr	Pro	Tyr	His	Leu	Phe	Ser	Ile	Trp	Glu	Leu	Thr	Ile	
				260					265					270			
	His	His	Asn	Ser	Tyr	Ser	His	His	Val	Met	Gln	Ala	Gly	Ile	Pro	Leu	
			275					280					285				
25	Ser	Thr	Gly	Leu	Ala	Phe	Leu	Asn	Ser	Cys	Leu	Asn	Pro	Ile	Leu	Tyr	
		290					295					300					
	Val	Leu	Ile	Ser	Lys	Lys	Phe	Gln	Ala	Arg	Phe	Arg	Ser	Ser	Val	Ala	
	305					310					315					320	
30	Glu	Ile	Leu	Lys	Tyr	Thr	Leu	Trp	Glu	Val	Ser	Cys	Ser	Gly	Thr	Val	
				325						330					335		
	Ser	Glu	Gln	Leu	Arg	Asn	Ser	Glu	Thr	Lys	Asn	Leu	Cys	Leu	Leu	Glu	
				340					345					350			
	Thr	Ala	Gln														
				355													

35 (166) INFORMATION FOR SEQ ID NO:165:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1089 base pairs

(B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:165:

	ATGGGCAACC ACACGTGGGA GGGCTGCCAC GTGGACTCGC GCGTGGACCA CCTCTTTCCG	60
	CCATCCCTCT ACATCTTTGT CATCGGCGTG GGGCTGCCCA CCAACTGCCT GGCTCTGTGG	120
	GCGGCCTACC GCCAGGTGCA ACAGCGCAAC GAGCTGGGCG TCTACCTGAT GAACCTCAGC	180
	ATCGCCGACC TGCTGTACAT CTGCACGCTG CCGCTGTGGG TGGACTACTT CCTGCACCAC	240
10	GACAACTGGA TCCACGGCCC CGGGTCTGTC AAGCTCTTTG GGTTCATCTT CTACACCAAT	300
	ATCTACATCA GCATCGCCTT CCTGTGCTGC ATCTCGGTGG ACCGCTACCT GGCTGTGGCC	360
	CACCCACTCC GCTTCGCCCC CCTGCGCCGC GTCAAGACCG CCGTGGCCGT GAGCTCCGTG	420
	GTCTGGGCCA CGGAGCTGGG CGCCAACCTG GCGCCCCTGT TCCATGACGA GCTCTTCCGA	480
	GACCGCTACA ACCACACCTT CTGCTTTGAG AAGTTCCCCA TGAAGGCTG GGTGGCCTGG	540
15	ATGAACCTCT ATCGGGTGTT CGTGGGCTTC CTCTTCCCGT GGGCGCTCAT GCTGCTGTGC	600
	TACCGGGGCA TCCTGCGGGC CGTGCGGGGC AGCGTGTCCT CCGAGCGCCA GGAGAAGGCC	660
	AAGATCGCGC GGCTGGCCCT CAGCCTCATC GCCATCGTGC TGGTCTGCTT TGCGCCCTAT	720
	CACGTGCTCT TGCTGTCCCG CAGCGCCATC TACCTGGGCC GCCCCTGGGA CTGCGGCTTC	780
	GAGGAGCGCG TCTTTTCTGC ATACCACAGC TCACTGGCTT TCACCAGCCT CAACTGTGTG	840
20	GCGGACCCCA TCCTCTACTG CCTGGTCAAC GAGGGCGCCC GCAGCGATGT GGCCAAGGCC	900
	CTGCACAACC TGCTCCGCTT TCTGGCCAGC GACAAGCCCC AGGAGATGGC CAATGCCTCG	960
	CTCACCCTGG AGACCCCACT CACCTCCAAG AGGAACAGCA CAGCCAAAGC CATGACTGGC	1020
	AGCTGGGCGG CCACTCCGCC TTCCCAGGGG GACCAGGTGC AGCTGAAGAT GCTGCCGCCA	1080
	GCACAATGA	1089

25 (167) INFORMATION FOR SEQ ID NO:166:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 362 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

30

(ii) MOLECULE TYPE: protein

115

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:166:

	Met	Gly	Asn	His	Thr	Trp	Glu	Gly	Cys	His	Val	Asp	Ser	Arg	Val	Asp	
	1				5					10					15		
5	His	Leu	Phe	Pro	Pro	Ser	Leu	Tyr	Ile	Phe	Val	Ile	Gly	Val	Gly	Leu	
				20				25					30				
	Pro	Thr	Asn	Cys	Leu	Ala	Leu	Trp	Ala	Ala	Tyr	Arg	Gln	Val	Gln	Gln	
			35				40					45					
	Arg	Asn	Glu	Leu	Gly	Val	Tyr	Leu	Met	Asn	Leu	Ser	Ile	Ala	Asp	Leu	
		50					55					60					
10	Leu	Tyr	Ile	Cys	Thr	Leu	Pro	Leu	Trp	Val	Asp	Tyr	Phe	Leu	His	His	
	65				70					75					80		
	Asp	Asn	Trp	Ile	His	Gly	Pro	Gly	Ser	Cys	Lys	Leu	Phe	Gly	Phe	Ile	
				85					90					95			
15	Phe	Tyr	Thr	Asn	Ile	Tyr	Ile	Ser	Ile	Ala	Phe	Leu	Cys	Cys	Ile	Ser	
				100				105					110				
	Val	Asp	Arg	Tyr	Leu	Ala	Val	Ala	His	Pro	Leu	Arg	Phe	Ala	Arg	Leu	
			115					120					125				
	Arg	Arg	Val	Lys	Thr	Ala	Val	Ala	Val	Ser	Ser	Val	Val	Trp	Ala	Thr	
		130					135					140					
20	Glu	Leu	Gly	Ala	Asn	Ser	Ala	Pro	Leu	Phe	His	Asp	Glu	Leu	Phe	Arg	
	145				150						155					160	
	Asp	Arg	Tyr	Asn	His	Thr	Phe	Cys	Phe	Glu	Lys	Phe	Pro	Met	Glu	Gly	
				165					170						175		
25	Trp	Val	Ala	Trp	Met	Asn	Leu	Tyr	Arg	Val	Phe	Val	Gly	Phe	Leu	Phe	
			180					185					190				
	Pro	Trp	Ala	Leu	Met	Leu	Leu	Ser	Tyr	Arg	Gly	Ile	Leu	Arg	Ala	Val	
			195				200					205					
	Arg	Gly	Ser	Val	Ser	Thr	Glu	Arg	Gln	Glu	Lys	Ala	Lys	Ile	Ala	Arg	
		210					215				220						
30	Leu	Ala	Leu	Ser	Leu	Ile	Ala	Ile	Val	Leu	Val	Cys	Phe	Ala	Pro	Tyr	
	225				230					235					240		
	His	Val	Leu	Leu	Leu	Ser	Arg	Ser	Ala	Ile	Tyr	Leu	Gly	Arg	Pro	Trp	
				245					250				255				
35	Asp	Cys	Gly	Phe	Glu	Glu	Arg	Val	Phe	Ser	Ala	Tyr	His	Ser	Ser	Leu	
			260					265					270				
	Ala	Phe	Thr	Ser	Leu	Asn	Cys	Val	Ala	Asp	Pro	Ile	Leu	Tyr	Cys	Leu	

275 280 285

Val Asn Glu Gly Ala Arg Ser Asp Val Ala Lys Ala Leu His Asn Leu
290 295 300

5 Leu Arg Phe Leu Ala Ser Asp Lys Pro Gln Glu Met Ala Asn Ala Ser
305 310 315 320

Leu Thr Leu Glu Thr Pro Leu Thr Ser Lys Arg Asn Ser Thr Ala Lys
325 330 335

Ala Met Thr Gly Ser Trp Ala Ala Thr Pro Pro Ser Gln Gly Asp Gln
340 345 350

10 Val Gln Leu Lys Met Leu Pro Pro Ala Gln
355 360

(168) INFORMATION FOR SEQ ID NO:167:

(i) SEQUENCE CHARACTERISTICS:

15 (A) LENGTH: 1002 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:167:

20 ATGGAGTCCT CAGGCAACCC AGAGAGCACC ACCTTTTTTT ACTATGACCT TCAGAGCCAG 60

CCGTGTGAGA ACCAGGCCTG GGTCTTTGCT ACCCTCGCCA CCACTGTCCT GTACTGCCTG 120

GTGTTTCTCC TCAGCCTAGT GGGCAACAGC CTGGTCCTGT GGGTCCTGGT GAAGTATGAG 180

AGCCTGGAGT CCCTCACCAA CATCTTCATC CTCAACCTGT GCCTCTCAGA CCTGGTGTTC 240

GCCTGCTTGT TGCCTGTGTG GATCTCCCCA TACCACTGGG GCTGGGTGCT GGGAGACTTC 300

25 CTCTGCAAAC TCCTCAATAT GATCTTCTCC ATCAGCCTCT ACAGCAGCAT CTTCTTCTCG 360

ACCATCATGA CCATCCACCG CTACCTGTCG GTAGTGAGCC CCCTCTCCAC CCTGCGCGTC 420

CCCACCCTCC GCTGCCGGGT GCTGGTGACC ATGGCTGTGT GGGTAGCCAG CATCCTGTCC 480

TCCATCCTCG ACACCATCTT CCACAAGGTG CTTTCTTCGG GCTGTGATTA TTCCGAATC 540

ACGTGGTACC TCACCTCCGT CTACCAGCAC AACCTCTTCT TCCTGCTGTC CCTGGGGATT 600

30 ATCCTGTTCT GCTACGTGGA GATCCTCAGG ACCCTGTTCC GCTCACGCTC CAAGCGGCGC 660

CACCGCACGA AAAAGCTCAT CTTCGCCATC GTGGTGGCCT ACTTCCTCAG CTGGGGTCCC 720

TACAACTTCA CCCTGTTTCT GCAGACGCTG TTTCGGACCC AGATCATCCG GAGCTGCGAG 780

117

GCCAAACAGC AGCTAGAATA CGCCCTGCTC ATCTGCCGCA ACCTCGCCTT CTCCCCTGCTC 840
 TGCTTTAACC CGGTGCTCTA TGTCTTCGTG GGGGTCAAGT TCCGCACACA CCTGAAACAT 900
 GTTCTCCGGC AGTTCTGGTT CTGCCGGCTG CAGGCACCCA GCCCAGCCTC GATCCCCCAG 960
 TCCCCTGGTG CCTTCGCCTA TGAGGGCGCC TCCTTCTACT GA 1002

5 (169) INFORMATION FOR SEQ ID NO:168:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 333 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:168:

Met	Glu	Ser	Ser	Gly	Asn	Pro	Glu	Ser	Thr	Thr	Phe	Phe	Tyr	Tyr	Asp	1	5	10	15
Leu	Gln	Ser	Gln	Pro	Cys	Glu	Asn	Gln	Ala	Trp	Val	Phe	Ala	Thr	Leu	20	25	30	
Ala	Thr	Thr	Val	Leu	Tyr	Cys	Leu	Val	Phe	Leu	Leu	Ser	Leu	Val	Gly	35	40	45	
Asn	Ser	Leu	Val	Leu	Trp	Val	Leu	Val	Lys	Tyr	Glu	Ser	Leu	Glu	Ser	50	55	60	
Leu	Thr	Asn	Ile	Phe	Ile	Leu	Asn	Leu	Cys	Leu	Ser	Asp	Leu	Val	Phe	65	70	75	80
Ala	Cys	Leu	Leu	Pro	Val	Trp	Ile	Ser	Pro	Tyr	His	Trp	Gly	Trp	Val	85	90	95	
Leu	Gly	Asp	Phe	Leu	Cys	Lys	Leu	Leu	Asn	Met	Ile	Phe	Ser	Ile	Ser	100	105	110	
Leu	Tyr	Ser	Ser	Ile	Phe	Phe	Leu	Thr	Ile	Met	Thr	Ile	His	Arg	Tyr	115	120	125	
Leu	Ser	Val	Val	Ser	Pro	Leu	Ser	Thr	Leu	Arg	Val	Pro	Thr	Leu	Arg	130	135	140	
Cys	Arg	Val	Leu	Val	Thr	Met	Ala	Val	Trp	Val	Ala	Ser	Ile	Leu	Ser	145	150	155	160
Ser	Ile	Leu	Asp	Thr	Ile	Phe	His	Lys	Val	Leu	Ser	Ser	Gly	Cys	Asp	165	170	175	
Tyr	Ser	Glu	Leu	Thr	Trp	Tyr	Leu	Thr	Ser	Val	Tyr	Gln	His	Asn	Leu	180	185	190	

Phe Phe Leu Leu Ser Leu Gly Ile Ile Leu Phe Cys Tyr Val Glu Ile
 195 200 205
 Leu Arg Thr Leu Phe Arg Ser Arg Ser Lys Arg Arg His Arg Thr Lys
 210 215 220
 5 Lys Leu Ile Phe Ala Ile Val Val Ala Tyr Phe Leu Ser Trp Gly Pro
 225 230 235 240
 Tyr Asn Phe Thr Leu Phe Leu Gln Thr Leu Phe Arg Thr Gln Ile Ile
 245 250 255
 10 Arg Ser Cys Glu Ala Lys Gln Gln Leu Glu Tyr Ala Leu Leu Ile Cys
 260 265 270
 Arg Asn Leu Ala Phe Ser His Cys Cys Phe Asn Pro Val Leu Tyr Val
 275 280 285
 Phe Val Gly Val Lys Phe Arg Thr His Leu Lys His Val Leu Arg Gln
 290 295 300
 15 Phe Trp Phe Cys Arg Leu Gln Ala Pro Ser Pro Ala Ser Ile Pro His
 305 310 315 320
 Ser Pro Gly Ala Phe Ala Tyr Glu Gly Ala Ser Phe Tyr
 325 330

(170) INFORMATION FOR SEQ ID NO:169:

- 20 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 987 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- 25 (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:169:

ATGGACAACG CCTCGTTCTC GGAGCCCTGG CCCGCCAACG CATCGGGCCC GGACCCGGCG 60
 CTGAGCTGCT CCAACGCGTC GACTCTGGCG CCGCTGCCGG CGCCGCTGGC GGTGGCTGTA 120
 CCAGTTGTCT ACGCGGTGAT CTGCGCCGTG GGTCTGGCGG GCAACTCCGC CGTGCTGTAC 180
 30 GTGTTGCTGC GGGCGCCCCG CATGAAGACC GTCACCAACC TGTTCATCCT CAACCTGGCC 240
 ATCGCCGACG AGCTCTTCAC GCTGGTGCTG CCCATCAACA TCGCCGACTT CCTGCTGCGG 300
 CAGTGGCCCT TCGGGGAGCT CATGTGCAAG CTCATCGTGG CTATCGACCA GTACAACACC 360
 TTCTCCAGCC TCTACTTCCT CACCGTCATG AGCGCCGACC GCTACCTGGT GGTGTTGGCC 420
 ACTGCGGAGT CGCGCCGGGT GGCCGGCCGC ACCTACAGCG CCGCGCGCGC GGTGAGCCTG 480

119

GCCGTGTGGG GGATCGTCAC ACTCGTCGTG CTGCCCTTCG CAGTCTTCGC CCGGCTAGAC 540
 GACGAGCAGG GCCGGCGCCA GTGCGTGCTA GTCTTTCCGC AGCCCGAGGC CTTCTGGTGG 600
 CGCGCGAGCC GCCTCTACAC GCTCGTGCTG GGCTTCGCCA TCCCCGTGTC CACCATCTGT 660
 GTCCTCTATA CCACCCTGCT GTGCCGGCTG CATGCCATGC GGCTGGACAG CCACGCCAAG 720
 5 GCCCTGGAGC GCGCCAAGAA GCGGGTGAAG TTCCTGGTGG TGGCAATCCT GCGGGTGTGC 780
 CTCCTCTGCT GGACGCCCTA CCACCTGAGC ACCGTGGTGG CGCTCACCAC CGACCTCCCG 840
 CAGACGCCGC TGGTCATCGC TATCTCCTAC TTCATCACCA GCCTGACGTA CGCCAACAGC 900
 TGCCTCAACC CCTTCCTCTA CGCCTTCTG GACGCCAGCT TCCGCAGGAA CCTCCGCCAG 960
 CTGATAACTT GCCGCGCGGC AGCCTGA 987

10 (171) INFORMATION FOR SEQ ID NO:170:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 328 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

15 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:170:

Met Asp Asn Ala Ser Phe Ser Glu Pro Trp Pro Ala Asn Ala Ser Gly
 1 5 10 15
 20 Pro Asp Pro Ala Leu Ser Cys Ser Asn Ala Ser Thr Leu Ala Pro Leu
 20 25 30
 Pro Ala Pro Leu Ala Val Ala Val Pro Val Val Tyr Ala Val Ile Cys
 35 40 45
 25 Ala Val Gly Leu Ala Gly Asn Ser Ala Val Leu Tyr Val Leu Leu Arg
 50 55 60
 Ala Pro Arg Met Lys Thr Val Thr Asn Leu Phe Ile Leu Asn Leu Ala
 65 70 75 80
 Ile Ala Asp Glu Leu Phe Thr Leu Val Leu Pro Ile Asn Ile Ala Asp
 85 90 95
 30 Phe Leu Leu Arg Gln Trp Pro Phe Gly Glu Leu Met Cys Lys Leu Ile
 100 105 110
 Val Ala Ile Asp Gln Tyr Asn Thr Phe Ser Ser Leu Tyr Phe Leu Thr
 115 120 125
 Val Met Ser Ala Asp Arg Tyr Leu Val Val Leu Ala Thr Ala Glu Ser

120

130 135 140

Arg Arg Val Ala Gly Arg Thr Tyr Ser Ala Ala Arg Ala Val Ser Leu
 145 150 155 160

Ala Val Trp Gly Ile Val Thr Leu Val Val Leu Pro Phe Ala Val Phe
 165 170 175

Ala Arg Leu Asp Asp Glu Gln Gly Arg Arg Gln Cys Val Leu Val Phe
 180 185 190

Pro Gln Pro Glu Ala Phe Trp Trp Arg Ala Ser Arg Leu Tyr Thr Leu
 195 200 205

Val Leu Gly Phe Ala Ile Pro Val Ser Thr Ile Cys Val Leu Tyr Thr
 210 215 220

Thr Leu Leu Cys Arg Leu His Ala Met Arg Leu Asp Ser His Ala Lys
 225 230 235 240

Ala Leu Glu Arg Ala Lys Lys Arg Val Lys Phe Leu Val Val Ala Ile
 245 250 255

Leu Ala Val Cys Leu Leu Cys Trp Thr Pro Tyr His Leu Ser Thr Val
 260 265 270

Val Ala Leu Thr Thr Asp Leu Pro Gln Thr Pro Leu Val Ile Ala Ile
 275 280 285

Ser Tyr Phe Ile Thr Ser Leu Thr Tyr Ala Asn Ser Cys Leu Asn Pro
 290 295 300

Phe Leu Tyr Ala Phe Leu Asp Ala Ser Phe Arg Arg Asn Leu Arg Gln
 305 310 315 320

Leu Ile Thr Cys Arg Ala Ala Ala
 325

(172) INFORMATION FOR SEQ ID NO:171:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1002 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:171:

ATGCAGGCCG CTGGGCACCC AGAGCCCCTT GACAGCAGGG GTCCTTCTC CCTCCCCACG 60

ATGGGTGCCA ACGTCTCTCA GGACAATGGC ACTGGCCACA ATGCCACCTT CTCCGAGCCA 120

CTGCCGTTCC TCTATGTGCT CCTGCCCGCC GTGTACTCCG GGATCTGTGC TGTGGGGCTG 180

ACTGGCAACA CGGCCGTCAT CCTTGTAATC CTAAGGGCGC CCAAGATGAA GACGGTGACC 240
 AACGTGTTCA TCCTGAACCT GGCCGTCGCC GACGGGGCTCT TCACGCTGGT ACTGCCTGTC 300
 AACATCGCGG AGCACCTGCT GCAGTACTGG CCCTTCGGGG AGCTGCTCTG CAAGCTGGTG 360
 CTGGCCGTCG ACCACTACAA CATCTTCTCC AGCATCTACT TCCTAGCCGT GATGAGCGTG 420
 5 GACCGATAACC TGGTGGTGCT GGCCACCGTG AGGTCCCGCC ACATGCCCTG GCGCACCTAC 480
 CGGGGGGCGA AGGTCGCCAG CCTGTGTGTC TGGCTGGGCG TCACGGTCCT GGTTCGCCCC 540
 TTCTTCTCTT TCGCTGGCGT CTACAGCAAC GAGCTGCAGG TCCAAGCTG TGGGCTGAGC 600
 TTCCCGTGGC CCGAGCAGGT CTGGTTCAAG GCCAGCCGTG TCTACACGTT GGTCTGGGC 660
 TTCGTGCTGC CCGTGTGCAC CATCTGTGTG CTCTACACAG ACCTCCTGCG CAGGCTGCGG 720
 10 GCCGTGCGGC TCCGCTCTGG AGCCAAGGCT CTAGGCAAGG CCAGGCGGAA GGTGAAAGTC 780
 CTGGTCCTCG TCGTGCTGGC CGTGTGCCTC CTCTGCTGGA CGCCCTTCCA CCTGGCCTCT 840
 GTCGTGGCCC TGACCACGGA CCTGCCCCAG ACCCACTGG TCATCAGTAT GTCCTACGTC 900
 ATCACCAGCC TCACGTACGC CAACTCGTGC CTGAACCCCT TCCTCTACGC CTTTCTAGAT 960
 GACAACTTCC GGAAGAACTT CCGCAGCATA TTGCGGTGCT GA 1002

15 (173) INFORMATION FOR SEQ ID NO:172:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 333 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - 20 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:172:

Met Gln Ala Ala Gly His Pro Glu Pro Leu Asp Ser Arg Gly Ser Phe
 1 5 10 15
 25 Ser Leu Pro Thr Met Gly Ala Asn Val Ser Gln Asp Asn Gly Thr Gly
 20 25 30
 His Asn Ala Thr Phe Ser Glu Pro Leu Pro Phe Leu Tyr Val Leu Leu
 35 40 45
 30 Pro Ala Val Tyr Ser Gly Ile Cys Ala Val Gly Leu Thr Gly Asn Thr
 50 55 60
 Ala Val Ile Leu Val Ile Leu Arg Ala Pro Lys Met Lys Thr Val Thr
 65 70 75 80

	Asn Val Phe Ile Leu Asn Leu Ala Val Ala Asp Gly Leu Phe Thr Leu	85	90	95
	Val Leu Pro Val Asn Ile Ala Glu His Leu Leu Gln Tyr Trp Pro Phe	100	105	110
5	Gly Glu Leu Leu Cys Lys Leu Val Leu Ala Val Asp His Tyr Asn Ile	115	120	125
	Phe Ser Ser Ile Tyr Phe Leu Ala Val Met Ser Val Asp Arg Tyr Leu	130	135	140
10	Val Val Leu Ala Thr Val Arg Ser Arg His Met Pro Trp Arg Thr Tyr	145	150	155
	Arg Gly Ala Lys Val Ala Ser Leu Cys Val Trp Leu Gly Val Thr Val	165	170	175
	Leu Val Leu Pro Phe Phe Ser Phe Ala Gly Val Tyr Ser Asn Glu Leu	180	185	190
15	Gln Val Pro Ser Cys Gly Leu Ser Phe Pro Trp Pro Glu Gln Val Trp	195	200	205
	Phe Lys Ala Ser Arg Val Tyr Thr Leu Val Leu Gly Phe Val Leu Pro	210	215	220
20	Val Cys Thr Ile Cys Val Leu Tyr Thr Asp Leu Leu Arg Arg Leu Arg	225	230	235
	Ala Val Arg Leu Arg Ser Gly Ala Lys Ala Leu Gly Lys Ala Arg Arg	245	250	255
	Lys Val Lys Val Leu Val Leu Val Val Leu Ala Val Cys Leu Leu Cys	260	265	270
25	Trp Thr Pro Phe His Leu Ala Ser Val Val Ala Leu Thr Thr Asp Leu	275	280	285
	Pro Gln Thr Pro Leu Val Ile Ser Met Ser Tyr Val Ile Thr Ser Leu	290	295	300
30	Thr Tyr Ala Asn Ser Cys Leu Asn Pro Phe Leu Tyr Ala Phe Leu Asp	305	310	315
	Asp Asn Phe Arg Lys Asn Phe Arg Ser Ile Leu Arg Cys	325	330	

(174) INFORMATION FOR SEQ ID NO:173:

- 35 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1107 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

123

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:173:

ATGGTCCTTG AGGTGAGTGA CCACCAAGTG CTAAATGACG CCGAGGTTGC CGCCCTCCTG 60
GAGAACTTCA GCTCTTCCTA TGA CTATGGA GAAAACGAGA GTGACTCGTG CTGTACCTCC 120
5 CCGCCCTGCC CACAGGACTT CAGCCTGAAC TTCGACCGGG CCTTCCTGCC AGCCCTCTAC 180
AGCCTCCTCT TTCTGCTGGG GCTGCTGGGC AACGGCGCGG TGGCAGCCGT GCTGCTGAGC 240
CGGCGGACAG CCCTGAGCAG CACCGACACC TTCCTGCTCC ACCTAGCTGT AGCAGACACG 300
CTGCTGGTGC TGACACTGCC GCTCTGGGCA GTGGACGCTG CCGTCCAGTG GGTCTTTGGC 360
TCTGGCCTCT GCAAAGTGGC AGGTGCCCTC TTCAACATCA ACTTCTACGC AGGAGCCCTC 420
10 CTGCTGGCCT GCATCAGCTT TGACCGCTAC CTGAACATAG TTCATGCCAC CCAGCTCTAC 480
CGCCGGGGGC CCCC GGCCCG CGTGACCCTC ACCTGCCTGG CTGTCTGGGG GCTCTGCCTG 540
CTTTTCGCCC TCCCAGACTT CATCTTCCTG TCGGCCCACC ACGACGAGCG CCTCAACGCC 600
ACCCACTGCC AATACAACTT CCCACAGGTG GGCCGCACGG CTCTGCGGGT GCTGCAGCTG 660
GTGGCTGGCT TTCTGCTGCC CCTGCTGGTC ATGGCCTACT GCTATGCCCA CATCCTGGCC 720
15 GTGCTGCTGG TTTCCAGGGG CCAGCGGCGC CTGCGGGCCA AGCGGCTGGT GGTGGTGGTC 780
GTGGTGGCCT TTGCCCTCTG CTGGACCCCC TATCACCTGG TGGTGCTGGT GGACATCCTC 840
ATGGACCTGG GCGCTTTGGC CCGCAACTGT GGCCGAGAAA GCAGGGTAGA CGTGGCCAAG 900
TCGGTCACCT CAGGCCTGGG CTACATGCAC TGCTGCCTCA ACCCGCTGCT CTATGCCTTT 960
GTAGGGGTCA AGTTCCGGGA GCGGATGTGG ATGCTGCTCT TGCGCCTGGG CTGCCCCAAC 1020
20 CAGAGAGGGC TCCAGAGGCA GCCATCGTCT TCCCGCCGGG ATTCATCCTG GTCTGAGACC 1080
TCAGAGGCCT CCTACTCGGG CTTGTGA 1107

(175) INFORMATION FOR SEQ ID NO:174:

(i) SEQUENCE CHARACTERISTICS:

- 25 (A) LENGTH: 368 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:174:

	Met	Val	Leu	Glu	Val	Ser	Asp	His	Gln	Val	Leu	Asn	Asp	Ala	Glu	Val	
	1				5					10					15		
	Ala	Ala	Leu	Leu	Glu	Asn	Phe	Ser	Ser	Ser	Tyr	Asp	Tyr	Gly	Glu	Asn	
				20					25					30			
5	Glu	Ser	Asp	Ser	Cys	Cys	Thr	Ser	Pro	Pro	Cys	Pro	Gln	Asp	Phe	Ser	
			35					40					45				
	Leu	Asn	Phe	Asp	Arg	Ala	Phe	Leu	Pro	Ala	Leu	Tyr	Ser	Leu	Leu	Phe	
		50					55					60					
10	Leu	Leu	Gly	Leu	Leu	Gly	Asn	Gly	Ala	Val	Ala	Ala	Val	Leu	Leu	Ser	
	65					70					75					80	
	Arg	Arg	Thr	Ala	Leu	Ser	Ser	Thr	Asp	Thr	Phe	Leu	Leu	His	Leu	Ala	
					85					90					95		
	Val	Ala	Asp	Thr	Leu	Leu	Val	Leu	Thr	Leu	Pro	Leu	Trp	Ala	Val	Asp	
				100					105					110			
15	Ala	Ala	Val	Gln	Trp	Val	Phe	Gly	Ser	Gly	Leu	Cys	Lys	Val	Ala	Gly	
			115					120					125				
	Ala	Leu	Phe	Asn	Ile	Asn	Phe	Tyr	Ala	Gly	Ala	Leu	Leu	Leu	Ala	Cys	
		130					135					140					
20	Ile	Ser	Phe	Asp	Arg	Tyr	Leu	Asn	Ile	Val	His	Ala	Thr	Gln	Leu	Tyr	
	145					150					155					160	
	Arg	Arg	Gly	Pro	Pro	Ala	Arg	Val	Thr	Leu	Thr	Cys	Leu	Ala	Val	Trp	
					165					170					175		
	Gly	Leu	Cys	Leu	Leu	Phe	Ala	Leu	Pro	Asp	Phe	Ile	Phe	Leu	Ser	Ala	
				180					185					190			
25	His	His	Asp	Glu	Arg	Leu	Asn	Ala	Thr	His	Cys	Gln	Tyr	Asn	Phe	Pro	
			195					200					205				
	Gln	Val	Gly	Arg	Thr	Ala	Leu	Arg	Val	Leu	Gln	Leu	Val	Ala	Gly	Phe	
		210					215					220					
30	Leu	Leu	Pro	Leu	Leu	Val	Met	Ala	Tyr	Cys	Tyr	Ala	His	Ile	Leu	Ala	
	225					230					235					240	
	Val	Leu	Leu	Val	Ser	Arg	Gly	Gln	Arg	Arg	Leu	Arg	Ala	Lys	Arg	Leu	
					245					250					255		
	Val	Val	Val	Val	Val	Val	Ala	Phe	Ala	Leu	Cys	Trp	Thr	Pro	Tyr	His	
				260					265					270			
35	Leu	Val	Val	Leu	Val	Asp	Ile	Leu	Met	Asp	Leu	Gly	Ala	Leu	Ala	Arg	
			275					280					285				
	Asn	Cys	Gly	Arg	Glu	Ser	Arg	Val	Asp	Val	Ala	Lys	Ser	Val	Thr	Ser	

125

290 295 300

Gly Leu Gly Tyr Met His Cys Cys Leu Asn Pro Leu Leu Tyr Ala Phe
305 310 315 320

5 Val Gly Val Lys Phe Arg Glu Arg Met Trp Met Leu Leu Leu Arg Leu
325 330 335

Gly Cys Pro Asn Gln Arg Gly Leu Gln Arg Gln Pro Ser Ser Ser Arg
340 345 350

Arg Asp Ser Ser Trp Ser Glu Thr Ser Glu Ala Ser Tyr Ser Gly Leu
355 360 365

10 (176) INFORMATION FOR SEQ ID NO:175:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1074 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 15 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:175:

ATGGCTGATG ACTATGGCTC TGAATCCACA TCTTCCATGG AAGACTACGT TAACTTCAAC 60

TTCAC TACTGACT TCTACTGTGA GAAAAACAAT GTCAGGCAGT TTGCGAGCCA TTCCTCCCA 120

20 CCCTTG TACT GGCTCGTGTT CATCGTGGGT GCCTTGGGCA ACAGTCTTGT TATCCTTGTC 180

TACTGG TACT GCACAAGAGT GAAGACCATG ACCGACATGT TCCTTTTGAA TTTGGCAATT 240

GCTGAC CTCC TCTTTCTTGT CACTCTTCCC TTCTGGGCCA TTGCTGCTGC TGACCACTGG 300

AAGTTCC AGA CCTTCATGTG CAAGGTGGTC AACAGCATGT ACAAGATGAA CTTCTACAGC 360

TGTGTG TTGC TGATCATGTG CATCAGCGTG GACAGGTACA TTGCCATTGC CCAGGCCATG 420

25 AGAGCAC ATA CTTGGAGGGA GAAAAGGCTT TTGTACAGCA AAATGGTTTG CTTTACCATC 480

TGGGTAT TGG CAGCTGCTCT CTGCATCCCA GAAATCTTAT ACAGCCAAAT CAAGGAGGAA 540

TCCGGCA TTG CTATCTGCAC CATGGTTTAC CCTAGCGATG AGAGCACCAA ACTGAAGTCA 600

GCTGTCT TGA CCCTGAAGGT CATTCTGGGG TTCTTCCTTC CCTTCGTGGT CATGGCTTGC 660

TGCTATA CCA TCATCATTCA CACCCTGATA CAAGCCAAGA AGTCTTCCAA GCACAAAGCC 720

30 AAGAAAG TGA CCATCACTGT CCTGACCGTC TTTGTCTTGT CTCAGTTTCC CTACAACTGC 780

ATTTTGT TGG TGCAGACCAT TGACGCCTAT GCCATGTTCA TCTCCAAC TGCCGTTTCC 840

ACCAACATTG ACATCTGCTT CCAGGTCACC CAGACCATCG CCTTCTTCCA CAGTTGCCTG 900
 AACCCGTGTTT TCTATGTTTT TGTGGGTGAG AGATTCCGCC GGGATCTCGT GAAAACCCTG 960
 AAGAACTTGG GTTGCATCAG CCAGGCCAG TGGGTTTCAT TTACAAGGAG AGAGGGAAGC 1020
 TTGAAGCTGT CGTCTATGTT GCTGGAGACA ACCTCAGGAG CACTCTCCCT CTGA 1074

5 (177) INFORMATION FOR SEQ ID NO:176:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 357 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

10 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:176:

Met Ala Asp Asp Tyr Gly Ser Glu Ser Thr Ser Ser Met Glu Asp Tyr
 1 5 10 15
 15 Val Asn Phe Asn Phe Thr Asp Phe Tyr Cys Glu Lys Asn Asn Val Arg
 20 25 30
 Gln Phe Ala Ser His Phe Leu Pro Pro Leu Tyr Trp Leu Val Phe Ile
 35 40 45
 20 Val Gly Ala Leu Gly Asn Ser Leu Val Ile Leu Val Tyr Trp Tyr Cys
 50 55 60
 Thr Arg Val Lys Thr Met Thr Asp Met Phe Leu Leu Asn Leu Ala Ile
 65 70 75 80
 Ala Asp Leu Leu Phe Leu Val Thr Leu Pro Phe Trp Ala Ile Ala Ala
 85 90 95
 25 Ala Asp Gln Trp Lys Phe Gln Thr Phe Met Cys Lys Val Val Asn Ser
 100 105 110
 Met Tyr Lys Met Asn Phe Tyr Ser Cys Val Leu Leu Ile Met Cys Ile
 115 120 125
 30 Ser Val Asp Arg Tyr Ile Ala Ile Ala Gln Ala Met Arg Ala His Thr
 130 135 140
 Trp Arg Glu Lys Arg Leu Leu Tyr Ser Lys Met Val Cys Phe Thr Ile
 145 150 155 160
 Trp Val Leu Ala Ala Ala Leu Cys Ile Pro Glu Ile Leu Tyr Ser Gln
 165 170 175
 35 Ile Lys Glu Glu Ser Gly Ile Ala Ile Cys Thr Met Val Tyr Pro Ser
 180 185 190

127

Asp Glu Ser Thr Lys Leu Lys Ser Ala Val Leu Thr Leu Lys Val Ile
 195 200 205
 Leu Gly Phe Phe Leu Pro Phe Val Val Met Ala Cys Cys Tyr Thr Ile
 210 215 220
 5 Ile Ile His Thr Leu Ile Gln Ala Lys Lys Ser Ser Lys His Lys Ala
 225 230 235 240
 Lys Lys Val Thr Ile Thr Val Leu Thr Val Phe Val Leu Ser Gln Phe
 245 250 255
 10 Pro Tyr Asn Cys Ile Leu Leu Val Gln Thr Ile Asp Ala Tyr Ala Met
 260 265 270
 Phe Ile Ser Asn Cys Ala Val Ser Thr Asn Ile Asp Ile Cys Phe Gln
 275 280 285
 Val Thr Gln Thr Ile Ala Phe Phe His Ser Cys Leu Asn Pro Val Leu
 290 295 300
 15 Tyr Val Phe Val Gly Glu Arg Phe Arg Arg Asp Leu Val Lys Thr Leu
 305 310 315 320
 Lys Asn Leu Gly Cys Ile Ser Gln Ala Gln Trp Val Ser Phe Thr Arg
 325 330 335
 20 Arg Glu Gly Ser Leu Lys Leu Ser Ser Met Leu Leu Glu Thr Thr Ser
 340 345 350
 Gly Ala Leu Ser Leu
 355

(178) INFORMATION FOR SEQ ID NO:177:

25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1110 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:177:

ATGGCCTCAT CGACCACTCG GGGCCCCAGG GTTTCTGACT TATTTTCTGG GCTGCCGCCG 60
 GCGGTCACAA CTCCCGCCAA CCAGAGCGCA GAGGCCTCGG CGGGCAACGG GTCGGTGGCT 120
 GGCGCGGACG CTCCAGCCGT CACGCCCTTC CAGAGCCTGC AGCTGGTGCA TCAGCTGAAG 180
 GGGCTGATCG TGCTGCTCTA CAGCGTCGTG GTGGTCGTGG GGCTGGTGGG CAACTGCCTG 240
 35 CTGGTGCTGG TGATCGCGCG GGTGCCGCGG CTGCACAACG TGACGAACTT CCTCATCGGC 300

AACCTGGCCT TGTCCGACGT GTCATGTGC ACCGCCTGCG TGCCGCTCAC GCTGGCCTAT 360
GCCTTCGAGC CACGCGGCTG GGTGTTCCGC GGC GGCCCTGT GCCACCTGGT CTTCTTCCTG 420
CAGCCGGTCA CCGTCTATGT GTCGGTGTTT ACGCTCACCA CCATCGCAGT GGACCGCTAC 480
GTCGTGCTGG TGCACCCGCT GAGGCGCGCA TCTCGCTGCG CCTCAGCCTA CGCTGTGCTG 540
5 GCCATCTGGG CGCTGTCCGC GGTGCTGGCG CTGCCGCCCC CCGTGCACAC CTATCACGTG 600
GAGCTCAAGC CGCACGACGT GCGCCTCTGC GAGGAGTTCT GGGGCTCCCA GGAGCGCCAG 660
CGCCAGCTCT ACGCCTGGGG GCTGCTGCTG GTCACCTACC TGCTCCCTCT GCTGGTCATC 720
CTCCTGTCTT ACGTCCGGGT GTCAGTGAAG CTCCGCAACC GCGTGGTGCC GGGCTGCGTG 780
ACCCAGAGCC AGGCCGACTG GGACCGCGCT CGGCGCCGGC GCACCAAATG CTTGCTGGTG 840
10 GTGGTCGTGG TGGTGTTCGC CGTCTGCTGG CTGCCGCTGC ACGTCTTCAA CCTGCTGCGG 900
GACCTCGACC CCCACGCCAT CGACCCTTAC GCCTTTGGGC TGGTGCAGCT GCTCTGCCAC 960
TGGCTCGCCA TGAGTTCGGC CTGCTACAAC CCCTTCATCT ACGCCTGGCT GCACGACAGC 1020
TTCCGCGAGG AGCTGCGCAA ACTGTTGGTC GCTTGGCCCC GCAAGATAGC CCCCCATGGC 1080
CAGAATATGA CCGTCAGCGT GGTCATCTGA 1110

15 (179) INFORMATION FOR SEQ ID NO:178:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 369 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant

20

(ii) MOLECULE TYPE: protein.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:178:

Met Ala Ser Ser Thr Thr Arg Gly Pro Arg Val Ser Asp Leu Phe Ser
1 5 10 15
25 Gly Leu Pro Pro Ala Val Thr Thr Pro Ala Asn Gln Ser Ala Glu Ala
 20 25 30
Ser Ala Gly Asn Gly Ser Val Ala Gly Ala Asp Ala Pro Ala Val Thr
 35 40 45
30 Pro Phe Gln Ser Leu Gln Leu Val His Gln Leu Lys Gly Leu Ile Val
 50 55 60
Leu Leu Tyr Ser Val Val Val Val Val Gly Leu Val Gly Asn Cys Leu
65 70 75 80

Leu	Val	Leu	Val	Ile	Ala	Arg	Val	Pro	Arg	Leu	His	Asn	Val	Thr	Asn	85	90	95
Phe	Leu	Ile	Gly	Asn	Leu	Ala	Leu	Ser	Asp	Val	Leu	Met	Cys	Thr	Ala	100	105	110
Cys	Val	Pro	Leu	Thr	Leu	Ala	Tyr	Ala	Phe	Glu	Pro	Arg	Gly	Trp	Val	115	120	125
Phe	Gly	Gly	Gly	Leu	Cys	His	Leu	Val	Phe	Phe	Leu	Gln	Pro	Val	Thr	130	135	140
Val	Tyr	Val	Ser	Val	Phe	Thr	Leu	Thr	Thr	Ile	Ala	Val	Asp	Arg	Tyr	145	150	155
Val	Val	Leu	Val	His	Pro	Leu	Arg	Arg	Ala	Ser	Arg	Cys	Ala	Ser	Ala	165	170	175
Tyr	Ala	Val	Leu	Ala	Ile	Trp	Ala	Leu	Ser	Ala	Val	Leu	Ala	Leu	Pro	180	185	190
Pro	Ala	Val	His	Thr	Tyr	His	Val	Glu	Leu	Lys	Pro	His	Asp	Val	Arg	195	200	205
Leu	Cys	Glu	Glu	Phe	Trp	Gly	Ser	Gln	Glu	Arg	Gln	Arg	Gln	Leu	Tyr	210	215	220
Ala	Trp	Gly	Leu	Leu	Leu	Val	Thr	Tyr	Leu	Leu	Pro	Leu	Leu	Val	Ile	225	230	235
Leu	Leu	Ser	Tyr	Val	Arg	Val	Ser	Val	Lys	Leu	Arg	Asn	Arg	Val	Val	245	250	255
Pro	Gly	Cys	Val	Thr	Gln	Ser	Gln	Ala	Asp	Trp	Asp	Arg	Ala	Arg	Arg	260	265	270
Arg	Arg	Thr	Lys	Cys	Leu	Leu	Val	Val	Val	Val	Val	Val	Phe	Ala	Val	275	280	285
Cys	Trp	Leu	Pro	Leu	His	Val	Phe	Asn	Leu	Leu	Arg	Asp	Leu	Asp	Pro	290	295	300
His	Ala	Ile	Asp	Pro	Tyr	Ala	Phe	Gly	Leu	Val	Gln	Leu	Leu	Cys	His	305	310	315
Trp	Leu	Ala	Met	Ser	Ser	Ala	Cys	Tyr	Asn	Pro	Phe	Ile	Tyr	Ala	Trp	325	330	335
Leu	His	Asp	Ser	Phe	Arg	Glu	Glu	Leu	Arg	Lys	Leu	Leu	Val	Ala	Trp	340	345	350
Pro	Arg	Lys	Ile	Ala	Pro	His	Gly	Gln	Asn	Met	Thr	Val	Ser	Val	Val	355	360	365

Ile

(180) INFORMATION FOR SEQ ID NO:179:

(i) SEQUENCE CHARACTERISTICS:

5

- (A) LENGTH: 1083 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:179:

```

10  ATGGACCCAG AAGAAACTTC AGTTTATTTG GATTATTACT ATGCTACGAG CCCAAACTCT    60
    GACATCAGGG AGACCCACTC CCATGTTCTT TACACCTCTG TCTTCCTTCC AGTCTTTTAC    120
    ACAGCTGTGT TCCTGACTGG AGTGCTGGGG AACCTTGTTT TCATGGGAGC GTTGCAATTC    180
    AAACCCGGCA GCCGAAGACT GATCGACATC TTTATCATCA ATCTGGCTGC CTCTGACTTC    240
    ATTTTCTCTG TCACATTGCC TCTCTGGGTG GATAAAGAAG CATCTCTAGG ACTGTGGAGG    300
15  ACGGGCTCCT TCCTGTGCAA AGGGAGCTCC TACATGATCT CCGTCAATAT GCACTGCAGT    360
    GTCTCCTGTC TCACTTGCAT GAGTGTGAC CGCTACCTGG CCATTGTGTG GCCAGTCGTA    420
    TCCAGGAAAT TCAGAAGGAC AGACTGTGCA TATGTAGTCT GTGCCAGCAT CTGGTTTATC    480
    TCCTGCCTGC TGGGGTTGCC TACTCTTCTG TCCAGGGAGC TCACGCTGAT TGATGATAAG    540
    CCATACTGTG CAGAGAAAAA GGCAACTCCA ATTAACTCA TATGGTCCCT GGTGGCCTTA    600
20  ATTTTCACCT TTTTGTCCC TTTGTTGAGC ATTGTGACCT GCTACTGTTG CATTGCAAGG    660
    AAGCTGTGTG CCCATTACCA GCAATCAGGA AAGCACAACA AAAAGCTGAA GAAATCTAAG    720
    AAGATCATCT TTATTGTCGT GGCAGCCTTT CTTGTCTCCT GGCTGCCCTT CAATACTTTC    780
    AAGTTCCTGG CCATTGTCTC TGGGTTGCGG CAAGAACACT ATTTACCCTC AGCTATTCTT    840
    CAGCTTG GTA TGGAGGTGAG TGGACCCTTG GCATTTGCCA ACAGCTGTGT CAACCCTTTC    900
25  ATTTACTATA TCTTCGACAG CTACATCCGC CGGGCCATTG TCCACTGCTT GTGCCCTTGC    960
    CTGAAAAACT ATGACTTTGG GAGTAGCACT GAGACATCAG ATAGTCACCT CACTAAGGCT   1020
    CTCTCCACCT TCATTCATGC AGAAGATTTT GCCAGGAGGA GGAAGAGGTC TGTGTCACTC   1080
    TAA                                                                    1083

```

(181) INFORMATION FOR SEQ ID NO:180:

30

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 360 amino acids

131

(B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:180:

	Met	Asp	Pro	Glu	Glu	Thr	Ser	Val	Tyr	Leu	Asp	Tyr	Tyr	Tyr	Ala	Thr	
	1				5					10					15		
	Ser	Pro	Asn	Ser	Asp	Ile	Arg	Glu	Thr	His	Ser	His	Val	Pro	Tyr	Thr	
			20						25					30			
10	Ser	Val	Phe	Leu	Pro	Val	Phe	Tyr	Thr	Ala	Val	Phe	Leu	Thr	Gly	Val	
			35						40					45			
	Leu	Gly	Asn	Leu	Val	Leu	Met	Gly	Ala	Leu	His	Phe	Lys	Pro	Gly	Ser	
		50					55					60					
15	Arg	Arg	Leu	Ile	Asp	Ile	Phe	Ile	Ile	Asn	Leu	Ala	Ala	Ser	Asp	Phe	
	65					70					75					80	
	Ile	Phe	Leu	Val	Thr	Leu	Pro	Leu	Trp	Val	Asp	Lys	Glu	Ala	Ser	Leu	
					85					90					95		
	Gly	Leu	Trp	Arg	Thr	Gly	Ser	Phe	Leu	Cys	Lys	Gly	Ser	Ser	Tyr	Met	
				100					105					110			
20	Ile	Ser	Val	Asn	Met	His	Cys	Ser	Val	Leu	Leu	Leu	Thr	Cys	Met	Ser	
			115					120					125				
	Val	Asp	Arg	Tyr	Leu	Ala	Ile	Val	Trp	Pro	Val	Val	Ser	Arg	Lys	Phe	
		130					135					140					
25	Arg	Arg	Thr	Asp	Cys	Ala	Tyr	Val	Val	Cys	Ala	Ser	Ile	Trp	Phe	Ile	
	145					150					155					160	
	Ser	Cys	Leu	Leu	Gly	Leu	Pro	Thr	Leu	Leu	Ser	Arg	Glu	Leu	Thr	Leu	
					165					170					175		
	Ile	Asp	Asp	Lys	Pro	Tyr	Cys	Ala	Glu	Lys	Lys	Ala	Thr	Pro	Ile	Lys	
				180						185					190		
30	Leu	Ile	Trp	Ser	Leu	Val	Ala	Leu	Ile	Phe	Thr	Phe	Phe	Val	Pro	Leu	
			195					200					205				
	Leu	Ser	Ile	Val	Thr	Cys	Tyr	Cys	Cys	Ile	Ala	Arg	Lys	Leu	Cys	Ala	
		210					215				220						
35	His	Tyr	Gln	Gln	Ser	Gly	Lys	His	Asn	Lys	Lys	Leu	Lys	Lys	Ser	Lys	
	225					230					235					240	
	Lys	Ile	Ile	Phe	Ile	Val	Val	Ala	Ala	Phe	Leu	Val	Ser	Trp	Leu	Pro	
					245					250					255		

Phe Asn Thr Phe Lys Phe Leu Ala Ile Val Ser Gly Leu Arg Gln Glu
 260 265 270

His Tyr Leu Pro Ser Ala Ile Leu Gln Leu Gly Met Glu Val Ser Gly
 275 280 285

5 Pro Leu Ala Phe Ala Asn Ser Cys Val Asn Pro Phe Ile Tyr Tyr Ile
 290 295 300

Phe Asp Ser Tyr Ile Arg Arg Ala Ile Val His Cys Leu Cys Pro Cys
 305 310 315 320

10 Leu Lys Asn Tyr Asp Phe Gly Ser Ser Thr Glu Thr Ser Asp Ser His
 325 330 335

Leu Thr Lys Ala Leu Ser Thr Phe Ile His Ala Glu Asp Phe Ala Arg
 340 345 350

Arg Arg Lys Arg Ser Val Ser Leu
 355 360

15 (182) INFORMATION FOR SEQ ID NO:181:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1020 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 20 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:181:

ATGAATGGCC TTGAAGTGGC TCCCCAGGT CTGATCACCA ACTTCTCCCT GGCCACGGCA 60
 GAGCAATGTG GCCAGGAGAC GCCACTGGAG AACATGCTGT TCGCCTCCTT CTACCTTCTG 120
 25 GATTTTATCC TGGCTTTAGT TGGCAATACC CTGGCTCTGT GGCTTTTCAT CCGAGACCAC 180
 AAGTCCGGGA CCCC GGCCAA CGTGTTCTCTG ATGCATCTGG CCGTGGCCGA CTTGTCGTGC 240
 GTGCTGGTCC TGCCACCCCG CCTGGTCTAC CACTTCTCTG GGAACCACTG GCCATTTGGG 300
 GAAATCGCAT GCCGTCTCAC CGGCTTCCTC TTCTACCTCA ACATGTACGC CAGCATCTAC 360
 TTCCTCACCT GCATCAGCGC CGACCGTTTC CTGGCCATTG TGCACCCGGT CAAGTCCCTC 420
 30 AAGCTCCGCA GGCCCTCTA CGCACACCTG GCCTGTGCCT TCCTGTGGGT GGTGGTGGCT 480
 GTGGCCATGG CCCC GCTGCT GGTGAGCCCA CAGACCGTGC AGACCAACCA CACGGTGGTC 540
 TGCCTGCAGC TGTACCGGGA GAAGGCCTCC CACCATGCCC TGGTGTCCCT GGCAGTGGCC 600
 TTCACCTTCC CGTTCATCAC CACGGTCACC TGCTACCTGC TGATCATCCG CAGCCTGCGG 660

133

CAGGGCCTGC GTGTGGAGAA GCGCCTCAAG ACCAAGGCAA AACGCATGAT CGCCATAGTG 720
 CTGGCCATCT TCCTGGTCTG CTTCTGTGCC TACCACGTCA ACCGCTCCGT CTACGTGCTG 780
 CACTACCGCA GCCATGGGGC CTCCTGCGCC ACCCAGCGCA TCCTGGCCCT GGCAAACCGC 840
 ATCACCTCCT GCCTCACCAG CCTCAACGGG GCACTCGACC CCATCATGTA TTTCTTCGTG 900
 5 GCTGAGAAGT TCCGCCACGC CCTGTGCAAC TTGCTCTGTG GCAAAAGGCT CAAGGGCCCG 960
 CCCCCAGCT TCGAAGGGAA AACCAACGAG AGCTCGCTGA GTGCCAAGTC AGAGCTGTGA 1020

(183) INFORMATION FOR SEQ ID NO:182:

(i) SEQUENCE CHARACTERISTICS:

10

- (A) LENGTH: 339 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:182:

15 Met Asn Gly Leu Glu Val Ala Pro Pro Gly Leu Ile Thr Asn Phe Ser
 1 5 10 15
 Leu Ala Thr Ala Glu Gln Cys Gly Gln Glu Thr Pro Leu Glu Asn Met
 20 20 25 30
 Leu Phe Ala Ser Phe Tyr Leu Leu Asp Phe Ile Leu Ala Leu Val Gly
 20 35 40 45
 Asn Thr Leu Ala Leu Trp Leu Phe Ile Arg Asp His Lys Ser Gly Thr
 50 55 60
 Pro Ala Asn Val Phe Leu Met His Leu Ala Val Ala Asp Leu Ser Cys
 65 70 75 80
 25 Val Leu Val Leu Pro Thr Arg Leu Val Tyr His Phe Ser Gly Asn His
 85 90 95
 Trp Pro Phe Gly Glu Ile Ala Cys Arg Leu Thr Gly Phe Leu Phe Tyr
 100 105 110
 Leu Asn Met Tyr Ala Ser Ile Tyr Phe Leu Thr Cys Ile Ser Ala Asp
 30 115 120 125
 Arg Phe Leu Ala Ile Val His Pro Val Lys Ser Leu Lys Leu Arg Arg
 130 135 140
 Pro Leu Tyr Ala His Leu Ala Cys Ala Phe Leu Trp Val Val Val Ala
 145 150 155 160
 35 Val Ala Met Ala Pro Leu Leu Val Ser Pro Gln Thr Val Gln Thr Asn

165 170 175
 His Thr Val Val Cys Leu Gln Leu Tyr Arg Glu Lys Ala Ser His His
 180 185 190
 5 Ala Leu Val Ser Leu Ala Val Ala Phe Thr Phe Pro Phe Ile Thr Thr
 195 200 205
 Val Thr Cys Tyr Leu Leu Ile Ile Arg Ser Leu Arg Gln Gly Leu Arg
 210 215 220
 Val Glu Lys Arg Leu Lys Thr Lys Ala Lys Arg Met Ile Ala Ile Val
 225 230 235 240
 10 Leu Ala Ile Phe Leu Val Cys Phe Val Pro Tyr His Val Asn Arg Ser
 245 250 255
 Val Tyr Val Leu His Tyr Arg Ser His Gly Ala Ser Cys Ala Thr Gln
 260 265 270
 15 Arg Ile Leu Ala Leu Ala Asn Arg Ile Thr Ser Cys Leu Thr Ser Leu
 275 280 285
 Asn Gly Ala Leu Asp Pro Ile Met Tyr Phe Phe Val Ala Glu Lys Phe
 290 295 300
 Arg His Ala Leu Cys Asn Leu Leu Cys Gly Lys Arg Leu Lys Gly Pro
 305 310 315 320
 20 Pro Pro Ser Phe Glu Gly Lys Thr Asn Glu Ser Ser Leu Ser Ala Lys
 325 330 335
 Ser Glu Leu

(183) INFORMATION FOR SEQ ID NO:183:

- 25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 996 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- 30 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:183:

ATGATCACCC TGAACAATCA AGATCAACCT GTCCCTTTTA ACAGCTCACA TCCAGATGAA 60
 TACAAAATTG CAGCCCTTGT CTTCTATAGC TGTATCTTCA TAATTGGATT ATTTGTTAAC 120
 ATCACTGCAT TATGGGTTTT CAGTTGTACC ACCAAGAAGA GAACCACGGT AACCATCTAT 180
 35 ATGATGAATG TGGCATTAGT GGAATTGATA TTTATAATGA CTTTACCCTT TCGAATGTTT 240

TATTATGCAA AAGATGAATG GCCATTGGA GAGTACTTCT GCCAGATTCT TGGAGCTCTC 300
 ACAGTGTTTT ACCCAAGCAT TGCTTTATGG CTTCTTGCCT TTATTAGTGC TGACAGATAC 360
 ATGGCCATTG TACAGCCGAA GTACGCCAAA GAACTTAAAA ACACGTGCAA AGCCGTGCTG 420
 GCGTGTGTGG GAGTCTGGAT AATGACCCTG ACCACGACCA CCCCTCTGCT ACTGCTCTAT 480
 5 AAAGACCCAG ATAAAGACTC CACTCCCGCC ACCTGCCTCA AGATTTCTGA CATCATCTAT 540
 CTAAAGCTG TGAACGTGCT GAACCTCACT CGACTGACAT TTTTTTCTT GATTCCTTTG 600
 TTCATCATGA TTGGGTGCTA CTTGGTCATT ATTCATAATC TCCTTCACGG CAGGACGTCT 660
 AAGCTGAAAC CCAAAGTCAA GGAGAAGTCC AAAAGGATCA TCATCACGCT GCTGGTGCAG 720
 GTGCTCGTCT GCTTTATGCC CTTCCACATC TGTTCGCTT TCCTGATGCT GGGAACGGGG 780
 10 GAGAATAGTT ACAATCCCTG GGGAGCCTTT ACCACCTTCC TCATGAACCT CAGCACGTGT 840
 CTGGATGTGA TTCTCTACTA CATCGTTTCA AAACAATTTT AGGCTCGAGT CATTAGTGTC 900
 ATGCTATACC GTAATTACCT TCGAAGCATG CGCAGAAAAA GTTTCCGATC TGGTAGTCTA 960
 AGGTCATAA GCAATATAAA CAGTGAAATG TTATGA 996

(185) INFORMATION FOR SEQ ID NO:184:

- 15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 331 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

- 20 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:184:

Met Ile Thr Leu Asn Asn Gln Asp Gln Pro Val Pro Phe Asn Ser Ser
 1 5 10 15

25 His Pro Asp Glu Tyr Lys Ile Ala Ala Leu Val Phe Tyr Ser Cys Ile
 20 25 30

Phe Ile Ile Gly Leu Phe Val Asn Ile Thr Ala Leu Trp Val Phe Ser
 35 40 45

Cys Thr Thr Lys Lys Arg Thr Thr Val Thr Ile Tyr Met Met Asn Val
 50 55 60

30 Ala Leu Val Asp Leu Ile Phe Ile Met Thr Leu Pro Phe Arg Met Phe
 65 70 75 80

Tyr Tyr Ala Lys Asp Glu Trp Pro Phe Gly Glu Tyr Phe Cys Gln Ile
 85 90 95

	Leu Gly Ala Leu Thr Val Phe Tyr Pro Ser Ile Ala Leu Trp Leu Leu	
	100	110
	Ala Phe Ile Ser Ala Asp Arg Tyr Met Ala Ile Val Gln Pro Lys Tyr	
	115	125
5	Ala Lys Glu Leu Lys Asn Thr Cys Lys Ala Val Leu Ala Cys Val Gly	
	130	140
	Val Trp Ile Met Thr Leu Thr Thr Thr Thr Pro Leu Leu Leu Leu Tyr	
	145	155
10	Lys Asp Pro Asp Lys Asp Ser Thr Pro Ala Thr Cys Leu Lys Ile Ser	
	165	175
	Asp Ile Ile Tyr Leu Lys Ala Val Asn Val Leu Asn Leu Thr Arg Leu	
	180	190
	Thr Phe Phe Phe Leu Ile Pro Leu Phe Ile Met Ile Gly Cys Tyr Leu	
	195	205
15	Val Ile Ile His Asn Leu Leu His Gly Arg Thr Ser Lys Leu Lys Pro	
	210	220
	Lys Val Lys Glu Lys Ser Lys Arg Ile Ile Ile Thr Leu Leu Val Gln	
	225	235
20	Val Leu Val Cys Phe Met Pro Phe His Ile Cys Phe Ala Phe Leu Met	
	245	255
	Leu Gly Thr Gly Glu Asn Ser Tyr Asn Pro Trp Gly Ala Phe Thr Thr	
	260	270
	Phe Leu Met Asn Leu Ser Thr Cys Leu Asp Val Ile Leu Tyr Tyr Ile	
	275	285
25	Val Ser Lys Gln Phe Gln Ala Arg Val Ile Ser Val Met Leu Tyr Arg	
	290	300
	Asn Tyr Leu Arg Ser Met Arg Arg Lys Ser Phe Arg Ser Gly Ser Leu	
	305	315
30	Arg Ser Leu Ser Asn Ile Asn Ser Glu Met Leu	
	325	330

(186) INFORMATION FOR SEQ ID NO:185:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1077 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:185:

ATGCCCTCTG TGTCTCCAGC GGGGCCCTCG GCCGGGGCAG TCCCCAATGC CACCGCAGTG 60
 ACAACAGTGC GGACCAATGC CAGCGGGCTG GAGGTGCCCC TGTTCACCT GTTTGCCCGG 120
 CTGGACGAGG AGCTGCATGG CACCTTCCCA GGCCTGTGCG TGGCGCTGAT GCGGGTGCAC 180
 5 GGAGCCATCT TCCTGGCAGG GCTGGTGCTC AACGGGCTGG CGCTGTACGT CTTCTGCTGC 240
 CGCACC CGG CCAAGACACC CTCAGTCATC TACACCATCA ACCTGGTGGT GACCGATCTA 300
 CTGGTAGGGC TGTCCCTGCC CACGCGCTTC GCTGTGTACT ACGGCGCCAG GGGCTGCCTG 360
 CGCTGTGCCT TCCCGCACGT CCTCGGTTAC TTCCTCAACA TGCACTGCTC CATCCTCTTC 420
 CTCACCTGCA TCTGCGTGGA CCGCTACCTG GCCATCGTGC GGCCCGAAGG CTCCCGCCGC 480
 10 TGCCGCCAGC CTGCCTGTGC CAGGGCCGTG TGCGCCTTCG TGTGGCTGGC CGCCGGTGCC 540
 GTCACCCTGT CGGTGCTGGG CGTGACAGGC AGCCGGCCCT GCTGCCGTGT CTTTGCGCTG 600
 ACTGTCCTGG AGTTCCTGCT GCCCTGCTG GTCATCAGCG TGTTTACCGG CCGCATCATG 660
 TGTGCACTGT CGCGGCCGGG TCTGCTCCAC CAGGGTCGCC AGCGCCGCGT GCGGGCCAAG 720
 CAGCTCCTGC TCACGGTGCT CATCATCTTT CTCGTCTGCT TCACGCCCTT CCACGCCCGC 780
 15 CAAGTGCCG TGGCGCTGTG GCCCGACATG CCACACCACA CGAGCCTCGT GGTCTACCAC 840
 GTGGCCGTGA CCCTCAGCAG CCTCAACAGC TGCATGGACC CCATCGTCTA CTGCTTCGTC 900
 ACCAGTGGCT TCCAGGCCAC CGTCCGAGGC CTCTTCGGCC AGCACGGAGA GCGTGAGCCC 960
 AGCAGCGGTG ACGTGGTCAG CATGCACAGG AGCTCCAAGG GCTCAGGCCG TCATCACATC 1020
 CTCAGTGCCG-GCCCTCACGC CCTCACCCAG GCCCTGGCTA ATGGGCCCCGA GGCTTAG 1077

20 (187) INFORMATION FOR SEQ ID NO:186:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 358 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 25 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:186:

Met Pro Ser Val Ser Pro Ala Gly Pro Ser Ala Gly Ala Val Pro Asn
 1 5 10 15
 30 Ala Thr Ala Val Thr Thr Val Arg Thr Asn Ala Ser Gly Leu Glu Val
 20 25 30

Pro Leu Phe His Leu Phe Ala Arg Leu Asp Glu Glu Leu His Gly Thr
 35 40 45
 Phe Pro Gly Leu Cys Val Ala Leu Met Ala Val His Gly Ala Ile Phe
 50 55 60
 5 Leu Ala Gly Leu Val Leu Asn Gly Leu Ala Leu Tyr Val Phe Cys Cys
 65 70 75 80
 Arg Thr Arg Ala Lys Thr Pro Ser Val Ile Tyr Thr Ile Asn Leu Val
 85 90 95
 10 Val Thr Asp Leu Leu Val Gly Leu Ser Leu Pro Thr Arg Phe Ala Val
 100 105 110
 Tyr Tyr Gly Ala Arg Gly Cys Leu Arg Cys Ala Phe Pro His Val Leu
 115 120 125
 Gly Tyr Phe Leu Asn Met His Cys Ser Ile Leu Phe Leu Thr Cys Ile
 130 135 140
 15 Cys Val Asp Arg Tyr Leu Ala Ile Val Arg Pro Glu Gly Ser Arg Ala
 145 150 155 160
 Cys Arg Gln Pro Ala Cys Ala Arg Ala Val Cys Ala Phe Val Trp Leu
 165 170 175
 20 Ala Ala Gly Ala Val Thr Leu Ser Val Leu Gly Val Thr Gly Ser Arg
 180 185 190
 Pro Cys Cys Arg Val Phe Ala Leu Thr Val Leu Glu Phe Leu Leu Pro
 195 200 205
 Leu Leu Val Ile Ser Val Phe Thr Gly Arg Ile Met Cys Ala Leu Ser
 210 215 220
 25 Arg Pro Gly Leu Leu His Gln Gly Arg Gln Arg Arg Val Arg Ala Lys
 225 230 235 240
 Gln Leu Leu Leu Thr Val Leu Ile Ile Phe Leu Val Cys Phe Thr Pro
 245 250 255
 30 Phe His Ala Arg Gln Val Ala Val Ala Leu Trp Pro Asp Met Pro His
 260 265 270
 His Thr Ser Leu Val Val Tyr His Val Ala Val Thr Leu Ser Ser Leu
 275 280 285
 Asn Ser Cys Met Asp Pro Ile Val Tyr Cys Phe Val Thr Ser Gly Phe
 290 295 300
 35 Gln Ala Thr Val Arg Gly Leu Phe Gly Gln His Gly Glu Arg Glu Pro
 305 310 315 320
 Ser Ser Gly Asp Val Val Ser Met His Arg Ser Ser Lys Gly Ser Gly

325 330 335
 Arg His His Ile Leu Ser Ala Gly Pro His Ala Leu Thr Gln Ala Leu
 340 345 350
 Ala Asn Gly Pro Glu Ala
 355

(188) INFORMATION FOR SEQ ID NO:187:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1050 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:187:

15 ATGAACTCCA CCTTGGATGG TAATCAGAGC AGCCACCCCTT TTTGCCTCTT GGCATTGTCG 60
 TATTTGGAAA CTGTCAATTT TTGCCTTTTG GAAGTATTGA TTATTGTCTT TCTAACTGTA 120
 TTGATTATTT CTGGCAACAT CATTGTGATT TTTGTATTTT ACTGTGCACC TTTGTTGAAC 180
 CATCACACTA CAAGTTATTT TATCCAGACT ATGGCATATG CTGACCTTTT TGTGTTGGGTG 240
 AGCTGCGTGG TCCCTTCTTT ATCACTCCTC CATCACCCCC TTCCAGTAGA GGAGTCCTTG 300
 ACTTGCCAGA TATTTGGTTT TGTAGTATCA GTTCTGAAGA GCGTCTCCAT GGCTTCTCTG 360
 20 GCCTGTATCA GCATTGATAG ATACATTGCC ATTACTAAAC CTTTAACCTA TAATACTCTG 420
 GTTACACCCT GGAGACTACG CCTGTGTATT TTCCTGATTT GGCTATACTC GACCCTGGTC 480
 TTCCTGCCTT CCTTTTTTCCA CTGGGGCAAA CCTGGATATC ATGGAGATGT GTTTCAGTGG 540
 TGTGCGGAGT CCTGGCACAC CGACTCCTAC TTCACCCTGT TCATCGTGAT GATGTTATAT 600
 GCCCCAGCAG CCCTTATTGT CTGCTTCACC TATTTCAACA TCTTCCGCAT CTGCCAACAG 660
 25 CACACAAAGG ATATCAGCGA AAGGCAAGCC CGCTTCAGCA GCCAGAGTGG GGAGACTGGG 720
 GAAGTGCAGG CCTGTCCTGA TAAGCGCTAT AAAATGGTCC TGTTTCGAAT CACTAGTGTA 780
 TTTTACATCC TCTGGTTGCC ATATATCATC TACTTCTTGT TGGAAAGCTC CACTGGCCAC 840
 AGCAACCGCT TCGCATCCTT CTTGACCACC TGGCTTGCTA TTAGTAACAG TTTCTGCAAC 900
 TGTGTAATTT ATAGTCTCTC CAACAGTGTA TTCAAAGAG GACTAAAGCG CCTCTCAGGG 960
 30 GCTATGTGTA CTTCTGTGC AAGTCAGACT ACAGCCAACG ACCCTTACAC AGTTAGAAGC 1020
 AAAGGCCCTC TTAATGGATG TCATATCTGA 1050

(189) INFORMATION FOR SEQ ID NO:188:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 349 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:188:

10 Met Asn Ser Thr Leu Asp Gly Asn Gln Ser Ser His Pro Phe Cys Leu
 1 5 10 15
 Leu Ala Phe Gly Tyr Leu Glu Thr Val Asn Phe Cys Leu Leu Glu Val
 20 25 30
 Leu Ile Ile Val Phe Leu Thr Val Leu Ile Ile Ser Gly Asn Ile Ile
 35 40 45
 15 Val Ile Phe Val Phe His Cys Ala Pro Leu Leu Asn His His Thr Thr
 50 55 60
 Ser Tyr Phe Ile Gln Thr Met Ala Tyr Ala Asp Leu Phe Val Gly Val
 65 70 75 80
 20 Ser Cys Val Val Pro Ser Leu Ser Leu Leu His His Pro Leu Pro Val
 85 90 95
 Glu Glu Ser Leu Thr Cys Gln Ile Phe Gly Phe Val Val Ser Val Leu
 100 105 110
 Lys Ser Val Ser Met Ala Ser Leu Ala Cys Ile Ser Ile Asp Arg Tyr
 115 120 125
 25 Ile Ala Ile Thr Lys Pro Leu Thr Tyr Asn Thr Leu Val Thr Pro Trp
 130 135 140
 Arg Leu Arg Leu Cys Ile Phe Leu Ile Trp Leu Tyr Ser Thr Leu Val
 145 150 155 160
 30 Phe Leu Pro Ser Phe Phe His Trp Gly Lys Pro Gly Tyr His Gly Asp
 165 170 175
 Val Phe Gln Trp Cys Ala Glu Ser Trp His Thr Asp Ser Tyr Phe Thr
 180 185 190
 Leu Phe Ile Val Met Met Leu Tyr Ala Pro Ala Ala Leu Ile Val Cys
 195 200 205
 35 Phe Thr Tyr Phe Asn Ile Phe Arg Ile Cys Gln Gln His Thr Lys Asp
 210 215 220

141

Ile Ser Glu Arg Gln Ala Arg Phe Ser Ser Gln Ser Gly Glu Thr Gly
 225 230 235 240

Glu Val Gln Ala Cys Pro Asp Lys Arg Tyr Lys Met Val Leu Phe Arg
 245 250 255

5 Ile Thr Ser Val Phe Tyr Ile Leu Trp Leu Pro Tyr Ile Ile Tyr Phe
 260 265 270

Leu Leu Glu Ser Ser Thr Gly His Ser Asn Arg Phe Ala Ser Phe Leu
 275 280 285

10 Thr Thr Trp Leu Ala Ile Ser Asn Ser Phe Cys Asn Cys Val Ile Tyr
 290 295 300

Ser Leu Ser Asn Ser Val Phe Gln Arg Gly Leu Lys Arg Leu Ser Gly
 305 310 315 320

Ala Met Cys Thr Ser Cys Ala Ser Gln Thr Thr Ala Asn Asp Pro Tyr
 325 330 335

15 Thr Val Arg Ser Lys Gly Pro Leu Asn Gly Cys His Ile
 340 345

(190) INFORMATION FOR SEQ ID NO:189:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1302 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:189:

25 ATGTGTTTTT CTCCATTCT GGAAATCAAC ATGCAGTCTG AATCTAACAT TACAGTGCGA 60
 GATGACATTG ATGACATCAA CACCAATATG TACCAACCAC TATCATATCC GTTAAGCTTT 120
 CAAGTGTCTC TCACCGGATT TCTTATGTTA GAAATTGTGT TGGGACTTGG CAGCAACCTC 180
 ACTGTATTGG TACTTTACTG CATGAAATCC AACTTAATCA ACTCTGTCAG TAACATTATT 240
 ACAATGAATC TTCATGTACT TGATGTAATA ATTTGTGTGG GATGTATTCC TCTAACTATA 300
 30 GTTATCCTTC TGCTTTCACT GGAGAGTAAC ACTGCTCTCA TTTGCTGTTT CCATGAGGCT 360
 TGTGTATCTT TTGCAAGTGT CTCAACAGCA ATCAACGTTT TTGCTATCAC TTTGGACAGA 420
 TATGACATCT CTGTAAACC TGCAAACCGA ATTCTGACAA TGGGCAGAGC TGTAATGTTA 480
 ATGATATCCA TTTGGATTTT TTCTTTTTTC TCTTTCCTGA TTCCTTTTAT TGAGGTAAAT 540
 TTTTTCAGTC TTCAAAGTGG AAATACCTGG GAAAACAAGA CACTTTTATG TGTCAGTACA 600

AATGAATACT AACTGAACT GGAATGTAT TATCACCTGT TAGTACAGAT CCCAATATTC 660
 TTTTTCAGTG TTGTAGTAAT GTTAATCACA TACACCAAAA TACTTCAGGC TCTTAATATT 720
 CGAATAGGCA CAAGATTTTC AACAGGGCAG AAGAAGAAAG CAAGAAAGAA AAAGACAATT 780
 TCTCTAACCA CACAACATGA GGCTACAGAC ATGTCACAAA GCAGTGGTGG GAGAAATGTA 840
 5 GTCTTTGGTG TAAGAACTTC AGTTTCTGTA ATAATTGCCC TCCGGCGAGC TGTGAAACGA 900
 CACCGTGAAC GACGAGAAAG ACAAAGAGA GTCAAGAGGA TGTCTTTATT GATTATTTCT 960
 ACATTTCTTC TCTGCTGGAC ACCAATTTCT GTTTTAAATA CCACCATTTT ATGTTTAGGC 1020
 CCAAGTGACC TTTTAGTAAA ATTAAGATTG TGTTTTTTAG TCATGGCTTA TGGAACAAC 1080
 ATATTTCAAC CTCTATTATA TGCATTCAC AGACAAAAAT TTCAAAGGT CTTGAAAAGT 1140
 10 AAAATGAAAA AGCGAGTTGT TTCTATAGTA GAAGCTGATC CCCTGCCTAA TAATGCTGTA 1200
 ATACACAAC CTTGGATAGA TCCCAAAAGA AACAAAAAAA TTACCTTTGA AGATAGTGAA 1260
 ATAAGAGAAA AACGTTTAGT GCCTCAGGTT GTCACAGACT AG 1302

(191) INFORMATION FOR SEQ ID NO:190:

- 15 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 433 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

- 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:190:

Met Cys Phe Ser Pro Ile Leu Glu Ile Asn Met Gln Ser Glu Ser Asn
 1 5 10 15

Ile Thr Val Arg Asp Asp Ile Asp Asp Ile Asn Thr Asn Met Tyr Gln
 20 25 30

25 Pro Leu Ser Tyr Pro Leu Ser Phe Gln Val Ser Leu Thr Gly Phe Leu
 35 40 45

Met Leu Glu Ile Val Leu Gly Leu Gly Ser Asn Leu Thr Val Leu Val
 50 55 60

30 Leu Tyr Cys Met Lys Ser Asn Leu Ile Asn Ser Val Ser Asn Ile Ile
 65 70 75 80

Thr Met Asn Leu His Val Leu Asp Val Ile Ile Cys Val Gly Cys Ile
 85 90 95

Pro Leu Thr Ile Val Ile Leu Leu Leu Ser Leu Glu Ser Asn Thr Ala

143

		100				105				110						
	Leu	Ile	Cys	Cys	Phe	His	Glu	Ala	Cys	Val	Ser	Phe	Ala	Ser	Val	Ser
			115					120					125			
5	Thr	Ala	Ile	Asn	Val	Phe	Ala	Ile	Thr	Leu	Asp	Arg	Tyr	Asp	Ile	Ser
		130					135					140				
	Val	Lys	Pro	Ala	Asn	Arg	Ile	Leu	Thr	Met	Gly	Arg	Ala	Val	Met	Leu
	145					150					155				160	
	Met	Ile	Ser	Ile	Trp	Ile	Phe	Ser	Phe	Phe	Ser	Phe	Leu	Ile	Pro	Phe
					165					170					175	
10	Ile	Glu	Val	Asn	Phe	Phe	Ser	Leu	Gln	Ser	Gly	Asn	Thr	Trp	Glu	Asn
				180					185					190		
	Lys	Thr	Leu	Leu	Cys	Val	Ser	Thr	Asn	Glu	Tyr	Tyr	Thr	Glu	Leu	Gly
			195					200					205			
15	Met	Tyr	Tyr	His	Leu	Leu	Val	Gln	Ile	Pro	Ile	Phe	Phe	Phe	Thr	Val
	210						215					220				
	Val	Val	Met	Leu	Ile	Thr	Tyr	Thr	Lys	Ile	Leu	Gln	Ala	Leu	Asn	Ile
	225					230					235				240	
	Arg	Ile	Gly	Thr	Arg	Phe	Ser	Thr	Gly	Gln	Lys	Lys	Lys	Ala	Arg	Lys
					245					250					255	
20	Lys	Lys	Thr	Ile	Ser	Leu	Thr	Thr	Gln	His	Glu	Ala	Thr	Asp	Met	Ser
				260					265					270		
	Gln	Ser	Ser	Gly	Gly	Arg	Asn	Val	Val	Phe	Gly	Val	Arg	Thr	Ser	Val
			275					280					285			
25	Ser	Val	Ile	Ile	Ala	Leu	Arg	Arg	Ala	Val	Lys	Arg	His	Arg	Glu	Arg
	290						295					300				
	Arg	Glu	Arg	Gln	Lys	Arg	Val	Lys	Arg	Met	Ser	Leu	Leu	Ile	Ile	Ser
	305				310					315					320	
	Thr	Phe	Leu	Leu	Cys	Trp	Thr	Pro	Ile	Ser	Val	Leu	Asn	Thr	Thr	Ile
				325					330						335	
30	Leu	Cys	Leu	Gly	Pro	Ser	Asp	Leu	Leu	Val	Lys	Leu	Arg	Leu	Cys	Phe
			340					345						350		
	Leu	Val	Met	Ala	Tyr	Gly	Thr	Thr	Ile	Phe	His	Pro	Leu	Leu	Tyr	Ala
			355					360					365			
35	Phe	Thr	Arg	Gln	Lys	Phe	Gln	Lys	Val	Leu	Lys	Ser	Lys	Met	Lys	Lys
	370						375					380				
	Arg	Val	Val	Ser	Ile	Val	Glu	Ala	Asp	Pro	Leu	Pro	Asn	Asn	Ala	Val
	385					390				395					400	

Ile His Asn Ser Trp Ile Asp Pro Lys Arg Asn Lys Lys Ile Thr Phe
 405 410 415

Glu Asp Ser Glu Ile Arg Glu Lys Arg Leu Val Pro Gln Val Val Thr
 420 425 430

5 Asp

(192) INFORMATION FOR SEQ ID NO:191:

(i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 1209 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:191:

15 ATGTTGTGTC CTTCCAAGAC AGATGGCTCA GGGCACTCTG GTAGGATTCA CCAGGAAACT 60
 CATGGAGAAG GGAAAAGGGA CAAGATTAGC AACAGTGAAG GGAGGGAGAA TGGTGGGAGA 120
 GGATTCCAGA TGAACGGTGG GTCGCTGGAG GCTGAGCATG CCAGCAGGAT GTCAGTTCTC 180
 AGAGCAAAGC CCATGTCAAA CAGCCAACGC TTGCTCCTTC TGTCCCCAGG ATCACCTCCT 240
 CGCACGGGGA GCATCTCCTA CATCAACATC ATCATGCCTT CGGTGTTCCG CACCATCTGC 300
 20 CTCCTGGGCA TCATCGGGAA CTCCACGGTC ATCTTCGCGG TCGTGAAGAA GTCCAAGCTG 360
 CACTGGTGCA ACAACGTCCC CGACATCTTC ATCATCAACC TCTCGGTAGT AGATCTCCTC 420
 TTTCTCCTGG GCATGCCCTT CATGATCCAC CAGCTCATGG GCAATGGGGT GTGGCACTTT 480
 GGGGAGACCA TGTGCACCCT CATCACGGCC ATGGATGCCA ATAGTCAGTT CACCAGCACC 540
 TACATCCTGA CCGCCATGGC CATTGACCGC TACCTGGCCA CTGTCCACCC CATCTCTTCC 600
 25 ACGAAGTTCC GGAAGCCCTC TGTGGCCACC CTGGTGATCT GCCTCCTGTG GGCCCTCTCC 660
 TTCATCAGCA TCACCCCTGT GTGGCTGTAT GCCAGACTCA TCCCCTTCCC AGGAGGTGCA 720
 GTGGGCTGCG GCATACGCCT GCCCAACCCA GACACTGACC TCTACTGGTT CACCCTGTAC 780
 CAGTTTTTCC TGGCCTTTGC CCTGCCTTTT GTGGTCATCA CAGCCGCATA CGTGAGGATC 840
 CTGCAGCGCA TGACGTCCTC AGTGGCCCCC GCCTCCCAGC GCAGCATCCG GCTGCGGACA 900
 30 AAGAGGGTGA AACGCACAGC CATCGCCATC TGTCTGGTCT TCTTTGTGTG CTGGGCACCC 960
 TACTATGTGC TACAGCTGAC CAGTTGTCC ATCAGCCGCC CGACCCTCAC CTTTGTCTAC 1020
 TTATACAATG CGGCCATCAG CTTGGGCTAT GCCAACAGCT GCCTCAACCC CTTTGTGTAC 1080

145

ATCGTGCTCT GTGAGACGTT CCGCAAACGC TTGGTCCTGT CGGTGAAGCC TGCAGCCCAG 1140
 GGGCAGCTTC GCGCTGTCAG CAACGCTCAG ACGGCTGACG AGGAGAGGAC AGAAAGCAAA 1200
 GGCACCTCA 1209

(193) INFORMATION FOR SEQ ID NO:192:

- 5 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 402 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

- 10 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:192:

Met Leu Cys Pro Ser Lys Thr Asp Gly Ser Gly His Ser Gly Arg Ile
 1 5 10 15
 His Gln Glu Thr His Gly Glu Gly Lys Arg Asp Lys Ile Ser Asn Ser
 15 20 25 30
 Glu Gly Arg Glu Asn Gly Gly Arg Gly Phe Gln Met Asn Gly Gly Ser
 35 40 45
 Leu Glu Ala Glu His Ala Ser Arg Met Ser Val Leu Arg Ala Lys Pro
 50 55 60
 Met Ser Asn Ser Gln Arg Leu Leu Leu Leu Ser Pro Gly Ser Pro Pro
 20 65 70 75 80
 Arg Thr Gly Ser Ile Ser Tyr Ile Asn Ile Ile Met Pro Ser Val Phe
 85 90 95
 Gly Thr Ile Cys Leu Leu Gly Ile Ile Gly Asn Ser Thr Val Ile Phe
 25 100 105 110
 Ala Val Val Lys Lys Ser Lys Leu His Trp Cys Asn Asn Val Pro Asp
 115 120 125
 Ile Phe Ile Ile Asn Leu Ser Val Val Asp Leu Leu Phe Leu Leu Gly
 130 135 140
 Met Pro Phe Met Ile His Gln Leu Met Gly Asn Gly Val Trp His Phe
 30 145 150 155 160
 Gly Glu Thr Met Cys Thr Leu Ile Thr Ala Met Asp Ala Asn Ser Gln
 165 170 175
 Phe Thr Ser Thr Tyr Ile Leu Thr Ala Met Ala Ile Asp Arg Tyr Leu
 35 180 185 190
 Ala Thr Val His Pro Ile Ser Ser Thr Lys Phe Arg Lys Pro Ser Val

		195		200		205
	Ala Thr Leu Val Ile Cys Leu Leu Trp Ala Leu Ser Phe Ile Ser Ile	210	215	220		
5	Thr Pro Val Trp Leu Tyr Ala Arg Leu Ile Pro Phe Pro Gly Gly Ala	225	230	235	240	
	Val Gly Cys Gly Ile Arg Leu Pro Asn Pro Asp Thr Asp Leu Tyr Trp	245	250	255		
	Phe Thr Leu Tyr Gln Phe Phe Leu Ala Phe Ala Leu Pro Phe Val Val	260	265	270		
10	Ile Thr Ala Ala Tyr Val Arg Ile Leu Gln Arg Met Thr Ser Ser Val	275	280	285		
	Ala Pro Ala Ser Gln Arg Ser Ile Arg Leu Arg Thr Lys Arg Val Lys	290	295	300		
15	Arg Thr Ala Ile Ala Ile Cys Leu Val Phe Phe Val Cys Trp Ala Pro	305	310	315	320	
	Tyr Tyr Val Leu Gln Leu Thr Gln Leu Ser Ile Ser Arg Pro Thr Leu	325	330	335		
	Thr Phe Val Tyr Leu Tyr Asn Ala Ala Ile Ser Leu Gly Tyr Ala Asn	340	345	350		
20	Ser Cys Leu Asn Pro Phe Val Tyr Ile Val Leu Cys Glu Thr Phe Arg	355	360	365		
	Lys Arg Leu Val Leu Ser Val Lys Pro Ala Ala Gln Gly Gln Leu Arg	370	375	380		
25	Ala Val Ser Asn Ala Gln Thr Ala Asp Glu Glu Arg Thr Glu Ser Lys	385	390	395	400	
	Gly Thr					

(194) INFORMATION FOR SEQ ID NO:193:

30 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1128 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:193:

ATGGATGTGA CTTCCCAAGC CCGGGGCGTG GGCCTGGAGA TGTACCCAGG CACCGCGCAC 60
GCTGCGGCCC CCAACACCAC CTCCCCGAG CTCAACCTGT CCCACCGCT CCTGGGCACC 120

147

GCCCTGGCCA ATGGGACAGG TGAGCTCTCG GAGCACCAGC AGTACGTGAT CGGCCTGTTC 180
 CTCTCGTGCC TCTACACCAT CTTCTCTTTC CCCATCGGCT TTGTGGGCAA CATCCTGATC 240
 CTGGTGGTGA ACATCAGCTT CCGCGAGAAG ATGACCATCC CCGACCTGTA CTTTCATCAAC 300
 CTGGCGGTGG CGGACCTCAT CCTGGTGGCC GACTCCCTCA TTGAGGTGTT CAACCTGCAC 360
 5 GAGCGGTACT ACGACATCGC CGTCCTGTGC ACCTTCATGT CGCTCTTCCT GCAGGTCAAC 420
 ATGTACAGCA GCGTCTTCTT CCTCACCTGG ATGAGCTTCG ACCGCTACAT CGCCCTGGCC 480
 AGGGCCATGC GCTGCAGCCT GTTCCGCACC AAGCACCACG CCCGGCTGAG CTGTGGCCTC 540
 ATCTGGATGG CATCCGTGTC AGCCACGCTG GTGCCCTTCA CCGCCGTGCA CCTGCAGCAC 600
 ACCGACGAGG CCTGCTTCTG TTTCGCGGAT GTCCGGGAGG TGCAGTGGCT CGAGGTCACG 660
 10 CTGGGCTTCA TCGTGCCCTT CGCCATCATC GGCCTGTGCT ACTCCCTCAT TGTCCGGGTG 720
 CTGGTCAGGG CGCACC GGCA CCGTGGGCTG CGGCCCCGGC GGCAGAAGGC GAAACGCATG 780
 ATCCTCGCGG TGGTGCTGGT CTTCTTCGTC TGCTGGCTGC CGGAGAACGT CTTTCATCAGC 840
 GTGCACCTCC TGCAGCGGAC GCAGCCTGGG GCCGCTCCCT GCAAGCAGTC TTTCCGCCAT 900
 GCCCACCCCC TCACGGGCCA CATTGTCAAC CTCGCCGCCT TCTCCAACAG CTGCCTAAAC 960
 15 CCCCTCATCT ACAGCTTTCT CGGGGAGACC TTCAGGGACA AGCTGAGGCT GTACATTGAG 1020
 CAGAAAACAA ATTTGCCGGC CCTGAACCGC TTCTGTACAG CTGCCCTGAA GGCCGTCATT 1080
 CCAGACAGCA CCGAGCAGTC GGATGTGAGG TTCAGCAGTG CCGTGTGA 1128

(195) INFORMATION FOR SEQ ID NO:194:

- 20 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 375 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:194:
 Met Asp Val Thr Ser Gln Ala Arg Gly Val Gly Leu Glu Met Tyr Pro
 1 5 10 15
 Gly Thr Ala His Ala Ala Ala Pro Asn Thr Thr Ser Pro Glu Leu Asn
 20 25 30
 30 Leu Ser His Pro Leu Leu Gly Thr Ala Leu Ala Asn Gly Thr Gly Glu
 35 40 45

Leu Ser Glu His Gln Gln Tyr Val Ile Gly Leu Phe Leu Ser Cys Leu
 50 55 60

Tyr Thr Ile Phe Leu Phe Pro Ile Gly Phe Val Gly Asn Ile Leu Ile
 65 70 75 80

5 Leu Val Val Asn Ile Ser Phe Arg Glu Lys Met Thr Ile Pro Asp Leu
 85 90 95

Tyr Phe Ile Asn Leu Ala Val Ala Asp Leu Ile Leu Val Ala Asp Ser
 100 105 110

10 Leu Ile Glu Val Phe Asn Leu His Glu Arg Tyr Tyr Asp Ile Ala Val
 115 120 125

Leu Cys Thr Phe Met Ser Leu Phe Leu Gln Val Asn Met Tyr Ser Ser
 130 135 140

Val Phe Phe Leu Thr Trp Met Ser Phe Asp Arg Tyr Ile Ala Leu Ala
 145 150 155 160

15 Arg Ala Met Arg Cys Ser Leu Phe Arg Thr Lys His His Ala Arg Leu
 165 170 175

Ser Cys Gly Leu Ile Trp Met Ala Ser Val Ser Ala Thr Leu Val Pro
 180 185 190

20 Phe Thr Ala Val His Leu Gln His Thr Asp Glu Ala Cys Phe Cys Phe
 195 200 205

Ala Asp Val Arg Glu Val Gln Trp Leu Glu Val Thr Leu Gly Phe Ile
 210 215 220

Val Pro Phe Ala Ile Ile Gly Leu Cys Tyr Ser Leu Ile Val Arg Val
 225 230 235 240

25 Leu Val Arg Ala His Arg His Arg Gly Leu Arg Pro Arg Arg Gln Lys
 245 250 255

Ala Lys Arg Met Ile Leu Ala Val Val Leu Val Phe Phe Val Cys Trp
 260 265 270

30 Leu Pro Glu Asn Val Phe Ile Ser Val His Leu Leu Gln Arg Thr Gln
 275 280 285

Pro Gly Ala Ala Pro Cys Lys Gln Ser Phe Arg His Ala His Pro Leu
 290 295 300

Thr Gly His Ile Val Asn Leu Ala Ala Phe Ser Asn Ser Cys Leu Asn
 305 310 315 320

35 Pro Leu Ile Tyr Ser Phe Leu Gly Glu Thr Phe Arg Asp Lys Leu Arg
 325 330 335

Leu Tyr Ile Glu Gln Lys Thr Asn Leu Pro Ala Leu Asn Arg Phe Cys

149

340 345 350

His Ala Ala Leu Lys Ala Val Ile Pro Asp Ser Thr Glu Gln Ser Asp
355 360 365

Val Arg Phe Ser Ser Ala Val
370 375

(196) INFORMATION FOR SEQ ID NO:195:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 960 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:195:

ATGCCATTCC CAAACTGCTC AGCCCCCAGC ACTGTGGTGG CCACAGCTGT GGGTGTCTTG 60

15 CTGGGGCTGG AGTGTGGGCT GGGTCTGCTG GGCAACGCGG TGGCGCTGTG GACCTTCCTG 120

TTCCGGGTCA GGGTGTGGAA GCCGTACGCT GTCTACCTGC TCAACCTGGC CCTGGCTGAC 180

CTGCTGTTGG CTGCGTGCCT GCCTTTCCTG GCCGCCTTCT ACCTGAGCCT CCAGGCTTGG 240

CATCTGGGCC GTGTGGGCTG CTGGGCCCTG CGCTTCCTGC TGGACCTCAG CCGCAGCGTG 300

GGGATGGCCT TCCTGGCCGC CGTGGCTTTG GACCGGTACC TCCGTGTGGT CCACCCTCGG 360

20 CTTAAGGTCA ACCTGCTGTC TCCTCAGGCG GCCCTGGGGG TCTCGGGCCT CGTCTGGCTC 420

CTGATGGTCG CCCTCACCTG CCCGGGCTTG CTCATCTCTG AGGCCGCCCA GAACTCCACC 480

AGGTGCCACA GTTTCCTACTC CAGGGCAGAC GGCTCCTTCA GCATCATCTG GCAGGAAGCA 540

CTCTCCTGCC TTCAGTTTGT CCTCCCCTTT GGCCTCATCG TGTTCTGCAA TGCAGGCATC 600

ATCAGGGCTC TCCAGAAAAG ACTCCGGGAG CCTGAGAAAC AGCCCAAGCT TCAGCGGGCC 660

25 AAGGCACTGG TCACCTTGGT GGTGGTGCTG TTTGCTCTGT GCTTTCCTGCC CTGCTTCCTG 720

GCCAGAGTCC TGATGCACAT CTTCCAGAAT CTGGGGAGCT GCAGGGCCCT TTGTGCAGTG 780

GCTCATACCT CGGATGTCAC GGGCAGCCTC ACCTACCTGC ACAGTGTCGT CAACCCCGTG 840

GTATACTGCT TCTCCAGCCC CACCTTCAGG AGCTCCTATC GGAGGGTCTT CCACACCCTC 900

CGAGGCAAAG GGCAGGCAGC AGAGCCCCCA GATTTC AACC CCAGAGACTC CTATTCCTGA 960

30 (197) INFORMATION FOR SEQ ID NO:196:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 319 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

5 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:196:

```

Met Pro Phe Pro Asn Cys Ser Ala Pro Ser Thr Val Val Ala Thr Ala
1           5           10           15
Val Gly Val Leu Leu Gly Leu Glu Cys Gly Leu Gly Leu Leu Gly Asn
10          20          25          30
Ala Val Ala Leu Trp Thr Phe Leu Phe Arg Val Arg Val Trp Lys Pro
          35          40          45
Tyr Ala Val Tyr Leu Leu Asn Leu Ala Leu Ala Asp Leu Leu Leu Ala
          50          55          60
15 Ala Cys Leu Pro Phe Leu Ala Ala Phe Tyr Leu Ser Leu Gln Ala Trp
          65          70          75          80
His Leu Gly Arg Val Gly Cys Trp Ala Leu Arg Phe Leu Leu Asp Leu
          85          90          95
20 Ser Arg Ser Val Gly Met Ala Phe Leu Ala Ala Val Ala Leu Asp Arg
          100          105          110
Tyr Leu Arg Val Val His Pro Arg Leu Lys Val Asn Leu Leu Ser Pro
          115          120          125
Gln Ala Ala Leu Gly Val Ser Gly Leu Val Trp Leu Leu Met Val Ala
          130          135          140
25 Leu Thr Cys Pro Gly Leu Leu Ile Ser Glu Ala Ala Gln Asn Ser Thr
          145          150          155          160
Arg Cys His Ser Phe Tyr Ser Arg Ala Asp Gly Ser Phe Ser Ile Ile
          165          170          175
30 Trp Gln Glu Ala Leu Ser Cys Leu Gln Phe Val Leu Pro Phe Gly Leu
          180          185          190
Ile Val Phe Cys Asn Ala Gly Ile Ile Arg Ala Leu Gln Lys Arg Leu
          195          200          205
Arg Glu Pro Glu Lys Gln Pro Lys Leu Gln Arg Ala Lys Ala Leu Val
          210          215          220
35 Thr Leu Val Val Val Leu Phe Ala Leu Cys Phe Leu Pro Cys Phe Leu
          225          230          235          240
Ala Arg Val Leu Met His Ile Phe Gln Asn Leu Gly Ser Cys Arg Ala

```

Leu His Ser Val Val Asn Pro Val Val Tyr Cys Phe Ser Ser Pro Thr
275 280 285

Gln Ala Ala Glu Pro Pro Asp Phe Asn Pro Arg Asp Ser Tyr Ser
305 310 315

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1143 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:197:

TACACAGACT GGAAATCCTC GGGGGCCCTC ATCCCTGCCA TCTACATGTT GGTCTTCCTC 120

AGGCGCTCAG CTGATATCTT CATTGCTAGC CTGGCGGTGG CTGACCTGAC CTTCGTGGTG 240

ACGCTGCCCC TGTGGGCTAC CTACACGTAC CGGGACTATG ACTGGCCCTT TGGGACCTTC 300

TTCTGCAAGC TCAGCAGCTA CCTCATCTTC GTCAACATGT ACGCCAGCGT CTTCTGCCTC 360

ACCGGCCTCA GCTTCGACCG CTACCTGGCC ATCGTGAGGC CAGTGGCCAA TGCTCGGCTG 420

GCCATGCCTG TCATGGTGTG ACGCACCACC GGGGACTTGG AGAACACCAC TAAGGTGCAG 540

TGCTACATGG ACTACTCCAT GGTGGCCACT GTGAGCTCAG AGTGGGCCTG GGAGGTGGGC 600

CTTGGGGTCT CGTCCACCAC CGTGGGCTTT GTGGTGCCCT TCACCATCAT GCTGACCTGT 660

TACTTCTTCA TCGCCCAAAC CATCGCTGGC CACTTCCGCA AGGAACGCAT CGAGGGCCTG 720

TGGATGCCCT ACCACCTGGT GAAGACGCTG TACATGCTGG GCAGCCTGCT GCACTGGCCC 840

TGTGACTTTG ACCTCTTCCT CATGAACATC TTCCCCTACT GCACCTGCAT CAGCTACGTC 900

AACAGCTGCC TCAACCCCTT CCTCTATGCC TTTTTCGACC CCCGCTTCCG CCAGGCCTGC 960
 ACCTCCATGC TCTGCTGTGG CCAGAGCAGG TGCGCAGGCA CCTCCCACAG CAGCAGTGGG 1020
 GAGAAGTCAG CCAGCTACTC TTCGGGGCAC AGCCAGGGGC CCGGCCCAA CATCGGCAAG 1080
 GGTGGAGAAC AGATGCACGA GAAATCCATC CCCTACAGCC AGGAGACCCT TGTGGTTGAC 1140
 5 TAG 1143

(199) INFORMATION FOR SEQ ID NO:198:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 380 amino acids
 (B) TYPE: amino acid
 10 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:198:

Met Glu Glu Gly Gly Asp Phe Asp Asn Tyr Tyr Gly Ala Asp Asn Gln
 1 5 10 15
 Ser Glu Cys Glu Tyr Thr Asp Trp Lys Ser Ser Gly Ala Leu Ile Pro
 20 25 30
 Ala Ile Tyr Met Leu Val Phe Leu Leu Gly Thr Thr Gly Asn Gly Leu
 35 40 45
 20 Val Leu Trp Thr Val Phe Arg Ser Ser Arg Glu Lys Arg Arg Ser Ala
 50 55 60
 Asp Ile Phe Ile Ala Ser Leu Ala Val Ala Asp Leu Thr Phe Val Val
 65 70 75 80
 25 Thr Leu Pro Leu Trp Ala Thr Tyr Thr Tyr Arg Asp Tyr Asp Trp Pro
 85 90 95
 Phe Gly Thr Phe Phe Cys Lys Leu Ser Ser Tyr Leu Ile Phe Val Asn
 100 105 110
 Met Tyr Ala Ser Val Phe Cys Leu Thr Gly Leu Ser Phe Asp Arg Tyr
 115 120 125
 30 Leu Ala Ile Val Arg Pro Val Ala Asn Ala Arg Leu Arg Leu Arg Val
 130 135 140
 Ser Gly Ala Val Ala Thr Ala Val Leu Trp Val Leu Ala Ala Leu Leu
 145 150 155 160
 35 Ala Met Pro Val Met Val Leu Arg Thr Thr Gly Asp Leu Glu Asn Thr
 165 170 175

153

Thr Lys Val Gln Cys Tyr Met Asp Tyr Ser Met Val Ala Thr Val Ser
 180 185 190
 Ser Glu Trp Ala Trp Glu Val Gly Leu Gly Val Ser Ser Thr Thr Val
 195 200 205
 5 Gly Phe Val Val Pro Phe Thr Ile Met Leu Thr Cys Tyr Phe Phe Ile
 210 215 220
 Ala Gln Thr Ile Ala Gly His Phe Arg Lys Glu Arg Ile Glu Gly Leu
 225 230 235 240
 10 Arg Lys Arg Arg Arg Leu Lys Ser Ile Ile Val Val Leu Val Val Thr
 245 250 255
 Phe Ala Leu Cys Trp Met Pro Tyr His Leu Val Lys Thr Leu Tyr Met
 260 265 270
 Leu Gly Ser Leu Leu His Trp Pro Cys Asp Phe Asp Leu Phe Leu Met
 275 280 285
 15 Asn Ile Phe Pro Tyr Cys Thr Cys Ile Ser Tyr Val Asn Ser Cys Leu
 290 295 300
 Asn Pro Phe Leu Tyr Ala Phe Phe Asp Pro Arg Phe Arg Gln Ala Cys
 305 310 315 320
 20 Thr Ser Met Leu Cys Cys Gly Gln Ser Arg Cys Ala Gly Thr Ser His
 325 330 335
 Ser Ser Ser Gly Glu Lys Ser Ala Ser Tyr Ser Ser Gly His Ser Gln
 340 345 350
 Gly Pro Gly Pro Asn Met Gly Lys Gly Gly Glu Gln Met His Glu Lys
 355 360 365
 25 Ser Ile Pro Tyr Ser Gln Glu Thr Leu Val Val Asp
 370 375 380

(200) INFORMATION FOR SEQ ID NO:199:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1119 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: DNA (genomic)
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:199:
 35 ATGAACTACC CGCTAACGCT GGAAATGGAC CTCGAGAACC TGGAGGACCT GTTCTGGGAA 60
 CTGGACAGAT TGGACAAC TAACGACACC TCCCTGGTGG AAAATCATCT CTGCCCTGCC 120

ACAGAGGGTC CCCTCATGGC CTCCTTCAAG GCCGTGTTTCG TGCCCGTGGC CTACAGCCTC 180
 ATCTTCCTCC TGGGCGTGAT CGGCAACGTC CTGGTGCTGG TGATCCTGGA GCGGCACCGG 240
 CAGACACGCA GTTCCACGGA GACCTTCCTG TTCCACCTGG CCGTGGCCGA CCTCCTGCTG 300
 GTCTTCATCT TGCCCTTTGC CGTGGCCGAG GGCTCTGTGG GCTGGGTCCT GGGGACCTTC 360
 5 CTCTGCAAAA CTGTGATTGC CCTGCACAAA GTCAACTTCT ACTGCAGCAG CCTGCTCCTG 420
 GCCTGCATCG CCGTGGACCG CTACCTGGCC ATTGTCCACG CCGTCCATGC CTACCGCCAC 480
 CGCCGCCTCC TCTCCATCCA CATCACCTGT GGGACCATCT GGCTGGTGGG CTTCCTCCTT 540
 GCCTTGCCAG AGATTCTCTT CGCCAAAGTC AGCCAAGGCC ATCACAACAA CTCCCTGCCA 600
 CGTTGCACCT TCTCCCAAGA GAACCAAGCA GAAACGCATG CCTGGTTCAC CTCCCGATTG 660
 10 CTCTACCATG TGGCGGGATT CCTGCTGCCC ATGCTGGTGA TGGGCTGGTG CTACGTGGGG 720
 GTAGTGACA GGTTCGCCA GGCCAGCGG CGCCCTCAGC GGCAGAAGGC AAAAAGGGTG 780
 GCCATCCTGG TGACAAGCAT CTTCTTCCTC TGCTGGTCAC CCTACCACAT CGTCATCTTC 840
 CTGGACACCC TGGCGAGGCT GAAGGCCGTG GACAATACCT GCAAGCTGAA TGGCTCTCTC 900
 CCCGTGGCCA TCACCATGTG TGAGTTCCTG GGCCTGGCCC ACTGCTGCCT CAACCCCATG 960
 15 CTCTACACTT TCGCCGGCGT GAAGTTCCGC AGTGACCTGT CGCGGCTCCT GACCAAGCTG 1020
 GGCTGTACCG GCCCTGCCTC CCTGTGCCAG CTCTTCCTA GCTGGCGCAG GAGCAGTCTC 1080
 TCTGAGTCAG AGAATGCCAC CTCTCTCACC ACGTTCTAG 1119

(201) INFORMATION FOR SEQ ID NO:200:

- 20 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 372 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant
 (ii) MOLECULE TYPE: protein

- 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:200:

Met Asn Tyr Pro Leu Thr Leu Glu Met Asp Leu Glu Asn Leu Glu Asp
 1 5 10 15

Leu Phe Trp Glu Leu Asp Arg Leu Asp Asn Tyr Asn Asp Thr Ser Leu
 20 25 30

30 Val Glu Asn His Leu Cys Pro Ala Thr Glu Gly Pro Leu Met Ala Ser
 35 40 45

Phe Lys Ala Val Phe Val Pro Val Ala Tyr Ser Leu Ile Phe Leu Leu

155

	50		55		60														
	Gly	Val	Ile	Gly	Asn	Val	Leu	Val	Leu	Val	Ile	Leu	Glu	Arg	His	Arg			
	65					70					75					80			
5	Gln	Thr	Arg	Ser	Ser	Thr	Glu	Thr	Phe	Leu	Phe	His	Leu	Ala	Val	Ala			
					85					90					95				
	Asp	Leu	Leu	Leu	Val	Phe	Ile	Leu	Pro	Phe	Ala	Val	Ala	Glu	Gly	Ser			
				100					105					110					
	Val	Gly	Trp	Val	Leu	Gly	Thr	Phe	Leu	Cys	Lys	Thr	Val	Ile	Ala	Leu			
			115					120					125						
10	His	Lys	Val	Asn	Phe	Tyr	Cys	Ser	Ser	Leu	Leu	Leu	Ala	Cys	Ile	Ala			
		130					135						140						
	Val	Asp	Arg	Tyr	Leu	Ala	Ile	Val	His	Ala	Val	His	Ala	Tyr	Arg	His			
	145					150					155					160			
15	Arg	Arg	Leu	Leu	Ser	Ile	His	Ile	Thr	Cys	Gly	Thr	Ile	Trp	Leu	Val			
					165					170					175				
	Gly	Phe	Leu	Leu	Ala	Leu	Pro	Glu	Ile	Leu	Phe	Ala	Lys	Val	Ser	Gln			
				180					185					190					
	Gly	His	His	Asn	Asn	Ser	Leu	Pro	Arg	Cys	Thr	Phe	Ser	Gln	Glu	Asn			
			195					200					205						
20	Gln	Ala	Glu	Thr	His	Ala	Trp	Phe	Thr	Ser	Arg	Phe	Leu	Tyr	His	Val			
		210					215					220							
	Ala	Gly	Phe	Leu	Leu	Pro	Met	Leu	Val	Met	Gly	Trp	Cys	Tyr	Val	Gly			
	225					230					235					240			
25	Val	Val	His	Arg	Leu	Arg	Gln	Ala	Gln	Arg	Arg	Pro	Gln	Arg	Gln	Lys			
					245					250					255				
	Ala	Lys	Arg	Val	Ala	Ile	Leu	Val	Thr	Ser	Ile	Phe	Phe	Leu	Cys	Trp			
				260					265					270					
	Ser	Pro	Tyr	His	Ile	Val	Ile	Phe	Leu	Asp	Thr	Leu	Ala	Arg	Leu	Lys			
			275					280					285						
30	Ala	Val	Asp	Asn	Thr	Cys	Lys	Leu	Asn	Gly	Ser	Leu	Pro	Val	Ala	Ile			
		290					295						300						
	Thr	Met	Cys	Glu	Phe	Leu	Gly	Leu	Ala	His	Cys	Cys	Leu	Asn	Pro	Met			
	305					310					315					320			
35	Leu	Tyr	Thr	Phe	Ala	Gly	Val	Lys	Phe	Arg	Ser	Asp	Leu	Ser	Arg	Leu			
				325					330						335				
	Leu	Thr	Lys	Leu	Gly	Cys	Thr	Gly	Pro	Ala	Ser	Leu	Cys	Gln	Leu	Phe			
				340					345					350					

Pro Ser Trp Arg Arg Ser Ser Leu Ser Glu Ser Glu Asn Ala Thr Ser
 355 360 365

Leu Thr Thr Phe
 370

5 (202) INFORMATION FOR SEQ ID NO:201:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1128 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

10 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:201:

ATGGATGTGA CTTCCAAGC CCGGGGCGTG GGCCTGGAGA TGTACCCAGG CACCGCGCAG 60
 CCTGCGGCCC CCAACACCAC CTCCCCGAG CTCAACCTGT CCCACCCGCT CCTGGGCACC 120
 15 GCCCTGGCCA ATGGGACAGG TGAGCTCTCG GAGCACCAGC AGTACGTGAT CGGCCTGTTC 180
 CTCTCGTGCC TCTACACCAT CTCCTCTTC CCCATCGGCT TTGTGGGCAA CATCCTGATC 240
 CTGGTGGTGA ACATCAGCTT CCGCGAGAAG ATGACCATCC CCGACCTGTA CTTTCATCAAC 300
 CTGGCGGTGG CGGACCTCAT CCTGGTGGCC GACTCCCTCA TTGAGGTGTT CAACCTGCAC 360
 GAGCGGTACT ACGACATCGC CGTCCTGTGC ACCTTCATGT CGCTCTTCCT GCAGGTCAAC 420
 20 ATGTACAGCA GCGTCTTCTT CCTCACCTGG ATGAGCTTCG ACCGCTACAT CGCCCTGGCC 480
 AGGGCCATGC GCTGCAGCCT GTTCCGCACC AAGCACCACG CCCGGCTGAG CTGTGGCCTC 540
 ATCTGGATGG CATCCGTGTC AGCCACGCTG GTGCCCTTCA CCGCCGTGCA CCTGCAGCAC 600
 ACCGACGAGG CCTGCTTCTG TTTGCGGAT GTCCGGGAGG TGCAGTGGCT CGAGGTCACG 660
 CTGGGCTTCA TCGTGCCCTT CGCCATCATC GGCCTGTGCT ACTCCCTCAT TGTCCGGGTG 720
 25 CTGGTCAGGG CGCACC GGCA CCGTGGGCTG CGGCCCCGGC GGCAGAAGGC GAAGCGCATG 780
 ATCCTCGCGG TGGTGCTGGT CTTCTTCGTC TGCTGGCTGC CGGAGAACGT CTTTCATCAGC 840
 GTGCACCTCC TGCAGCGGAC GCAGCCTGGG GCCGCTCCCT GCAAGCAGTC TTTCCGCCAT 900
 GCCCACCCCC TCACGGGCCA CATGTGTAAC CTCACCGCCT TCTCCAACAG CTGCCTAAAC 960
 CCCCTCATCT ACAGCTTTCT CGGGGAGACC TTCAGGGACA AGCTGAGGCT GTACATTGAG 1020
 30 CAGAAAACAA ATTTGCCGGC CCTGAACCGC TTCTGTCACG CTGCCCTGAA GGCCGTGATT 1080
 CCAGACAGCA CCGAGCAGTC GGATGTGAGG TTCAGCAGTG CCGTGTAG 1128

157

(203) INFORMATION FOR SEQ ID NO:202:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 375 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:202:

5 Met Asp Val Thr Ser Gln Ala Arg Gly Val Gly Leu Glu Met Tyr Pro
 1 5 10 15
 Gly Thr Ala Gln Pro Ala Ala Pro Asn Thr Thr Ser Pro Glu Leu Asn
 20 25 30
 Leu Ser His Pro Leu Leu Gly Thr Ala Leu Ala Asn Gly Thr Gly Glu
 35 40 45
 15 Leu Ser Glu His Gln Gln Tyr Val Ile Gly Leu Phe Leu Ser Cys Leu
 50 55 60
 Tyr Thr Ile Phe Leu Phe Pro Ile Gly Phe Val Gly Asn Ile Leu Ile
 65 70 75 80
 20 Leu Val Val Asn Ile Ser Phe Arg Glu Lys Met Thr Ile Pro Asp Leu
 85 90 95
 Tyr Phe Ile Asn Leu Ala Val Ala Asp Leu Ile Leu Val Ala Asp Ser
 100 105 110
 Leu Ile Glu Val Phe Asn Leu His Glu Arg Tyr Tyr Asp Ile Ala Val
 115 120 125
 25 Leu Cys Thr Phe Met Ser Leu Phe Leu Gln Val Asn Met Tyr Ser Ser
 130 135 140
 Val Phe Phe Leu Thr Trp Met Ser Phe Asp Arg Tyr Ile Ala Leu Ala
 145 150 155 160
 30 Arg Ala Met Arg Cys Ser Leu Phe Arg Thr Lys His His Ala Arg Leu
 165 170 175
 Ser Cys Gly Leu Ile Trp Met Ala Ser Val Ser Ala Thr Leu Val Pro
 180 185 190
 Phe Thr Ala Val His Leu Gln His Thr Asp Glu Ala Cys Phe Cys Phe
 195 200 205
 35 Ala Asp Val Arg Glu Val Gln Trp Leu Glu Val Thr Leu Gly Phe Ile
 210 215 220
 Val Pro Phe Ala Ile Ile Gly Leu Cys Tyr Ser Leu Ile Val Arg Val

225 230 235 240
 Leu Val Arg Ala His Arg His Arg Gly Leu Arg Pro Arg Arg Gln Lys
 245 250 255
 5 Ala Lys Arg Met Ile Leu Ala Val Val Leu Val Phe Phe Val Cys Trp
 260 265 270
 Leu Pro Glu Asn Val Phe Ile Ser Val His Leu Leu Gln Arg Thr Gln
 275 280 285
 Pro Gly Ala Ala Pro Cys Lys Gln Ser Phe Arg His Ala His Pro Leu
 290 295 300
 10 Thr Gly His Ile Val Asn Leu Thr Ala Phe Ser Asn Ser Cys Leu Asn
 305 310 315 320
 Pro Leu Ile Tyr Ser Phe Leu Gly Glu Thr Phe Arg Asp Lys Leu Arg
 325 330 335
 15 Leu Tyr Ile Glu Gln Lys Thr Asn Leu Pro Ala Leu Asn Arg Phe Cys
 340 345 350
 His Ala Ala Leu Lys Ala Val Ile Pro Asp Ser Thr Glu Gln Ser Asp
 355 360 365
 Val Arg Phe Ser Ser Ala Val
 370 375

20 (204) INFORMATION FOR SEQ ID NO:203:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1137 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - 25 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:203:

ATGGACCTGG GGAAACCAAT GAAAAGCGTG CTGGTGGTGG CTCTCCTTGT CATTTTCCAG 60
 GTATGCCTGT GTCAAGATGA GGTACGGGAC GATTACATCG GAGACAACAC CACAGTGGAC 120
 30 TACACTTTGT TCGAGTCTTT GTGCTCCAAG AAGGACGTGC GGAAC TTAA AGCCTGGTTC 180
 CTCCCTATCA TGTACTCCAT CATTTGTTTC GTGGGCCTAC TGGGCAATGG GCTGGTCGTG 240
 TTGACCTATA TCTATTTCAA GAGGCTCAAG ACCATGACCG ATACCTACCT GCTCAACCTG 300
 GCGGTGGCAG ACATCCTCTT CCTCCTGACC CTTCCCTTCT GGGCCTACAG CGCGCCAAG 360
 TCCTGGGTCT TCGGTGTCCA CTTTGTGCAAG CTCATCTTTG CCATCTACAA GATGAGCTTC 420

159

TTCAGTGGCA TGCTCCTACT TCTTTGCATC AGCATTGACC GCTACGTGGC CÁTCTGCCAG 480
 GCTGTCTCAG CTCACCGCCA CCGTGCCCCG GTCCTTCTCA TCAGCAAGCT GTCCTGTGTG 540
 GGCATCTGGA TACTAGCCAC AGTGCTCTCC ATCCCAAGAGC TCCTGTACAG TGACCTCCAG 600
 AGGAGCAGCA GTGAGCAAGC GATGCGATGC TCTCTCATCA CAGAGCATGT GGAGGCCTTT 660
 5 ATCACCATCC AGGTGGCCCA GATGGTGATC GGCTTTCTGG TCCCCCTGCT GGCCATGAGC 720
 TTCTGTTACC TTGTCATCAT CCGCACCCCTG CTCCAGGCAC GCAACTTTGA GCGCAACAAG 780
 GCCAAAAAGG TGATCATCGC TGTGGTCGTG GTCTTCATAG TCTTCCAGCT GCCCTACAAT 840
 GGGGTGGTCC TGGCCCAGAC GGTGGCCAAC TTCAACATCA CCAGTAGCAC CTGTGAGCTC 900
 AGTAAGCAAC TCAACATCGC CTACGACGTC ACCTACAGCC TGGCCTGCGT CCGCTGCTGC 960
 10 GTCAACCCTT TCTTGACGC CTTTCATCGG GTCAAGTTCC GCAACGATCT CTTCAAGCTC 1020
 TTCAAGGACC TGGGCTGCCT CAGCCAGGAG CAGCTCCGGC AGTGGTCTTC CTGTGGGCAC 1080
 ATCCGGCGCT CCTCCATGAG TGTGGAGGCC GAGACCACCA CCACCTTCTC CCCATAG 1137

(205) INFORMATION FOR SEQ ID NO:204:

15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 378 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:204:

Met Asp Leu Gly Lys Pro Met Lys Ser Val Leu Val Val Ala Leu Leu
 1 5 10 15

Val Ile Phe Gln Val Cys Leu Cys Gln Asp Glu Val Thr Asp Asp Tyr
 20 25 30

25 Ile Gly Asp Asn Thr Thr Val Asp Tyr Thr Leu Phe Glu Ser Leu Cys
 35 40 45

Ser Lys Lys Asp Val Arg Asn Phe Lys Ala Trp Phe Leu Pro Ile Met
 50 55 60

30 Tyr Ser Ile Ile Cys Phe Val Gly Leu Leu Gly Asn Gly Leu Val Val
 65 70 75 80

Leu Thr Tyr Ile Tyr Phe Lys Arg Leu Lys Thr Met Thr Asp Thr Tyr
 85 90 95

Leu Leu Asn Leu Ala Val Ala Asp Ile Leu Phe Leu Leu Thr Leu Pro

	100	105	110
	Phe Trp Ala Tyr Ser Ala Ala Lys Ser Trp Val Phe Gly Val His Phe		
	115	120	125
5	Cys Lys Leu Ile Phe Ala Ile Tyr Lys Met Ser Phe Phe Ser Gly Met		
	130	135	140
	Leu Leu Leu Leu Cys Ile Ser Ile Asp Arg Tyr Val Ala Ile Val Gln		
	145	150	155
	Ala Val Ser Ala His Arg His Arg Ala Arg Val Leu Leu Ile Ser Lys		
		165	170
10	Leu Ser Cys Val Gly Ile Trp Ile Leu Ala Thr Val Leu Ser Ile Pro		
		180	185
	Glu Leu Leu Tyr Ser Asp Leu Gln Arg Ser Ser Ser Glu Gln Ala Met		
		195	200
15	Arg Cys Ser Leu Ile Thr Glu His Val Glu Ala Phe Ile Thr Ile Gln		
		210	215
	Val Ala Gln Met Val Ile Gly Phe Leu Val Pro Leu Leu Ala Met Ser		
		225	230
	Phe Cys Tyr Leu Val Ile Ile Arg Thr Leu Leu Gln Ala Arg Asn Phe		
		245	250
20	Glu Arg Asn Lys Ala Lys Lys Val Ile Ile Ala Val Val Val Val Phe		
		260	265
	Ile Val Phe Gln Leu Pro Tyr Asn Gly Val Val Leu Ala Gln Thr Val		
		275	280
25	Ala Asn Phe Asn Ile Thr Ser Ser Thr Cys Glu Leu Ser Lys Gln Leu		
		290	295
	Asn Ile Ala Tyr Asp Val Thr Tyr Ser Leu Ala Cys Val Arg Cys Cys		
		305	310
	Val Asn Pro Phe Leu Tyr Ala Phe Ile Gly Val Lys Phe Arg Asn Asp		
		325	330
30	Leu Phe Lys Leu Phe Lys Asp Leu Gly Cys Leu Ser Gln Glu Gln Leu		
		340	345
	Arg Gln Trp Ser Ser Cys Arg His Ile Arg Arg Ser Ser Met Ser Val		
		355	360
35	Glu Ala Glu Thr Thr Thr Thr Phe Ser Pro		
		370	375

(206) INFORMATION FOR SEQ ID NO:205:

161

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1086 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:205:

ATGGATATAC AAATGGCAAA CAATTTTACT CCGCCCTCTG CAACTCCTCA GGGAAATGAC 60
TGTGACCTCT ATGCACATCA CAGCACGGCC AGGATAGTAA TGCCTCTGCA TTACAGCCTC 120
10 GTCTTCATCA TTGGGCTCGT GGGAAACTTA CTAGCCTTGG TCGTCATTGT TCAAAACAGG 180
AAAAAAATCA ACTCTACCAC CCTCTATTCA ACAAATTTGG TGATTTCTGA TATACTTTTT 240
ACCACGGCTT TGCCTACACG AATAGCCTAC TATGCAATGG GCTTTGACTG GAGAATCGGA 300
GATGCCTTGT GTAGGATAAC TGCGCTAGTG TTTTACATCA ACACATATGC AGGTGTGAAC 360
TTTATGACCT GCCTGAGTAT TGACCGCTTC ATTGCTGTGG TGCACCCTCT ACGCTACAAC 420
15 AAGATAAAAA GGATTGAACA TGCAAAGGC GTGTGCATAT TTGTCTGGAT TCTAGTATTT 480
GCTCAGACAC TCCCCTCCT CATCAACCCT ATGTCAAAGC AGGAGGCTGA AAGGATTACA 540
TGCATGGAGT ATCCAACTT TGAAGAACT AAATCTCTTC CCTGGATTCT GCTTGGGGCA 600
TGTTTCATAG GATATGTACT TCCACTTATA ATCATTCTCA TCTGCTATTC TCAGATCTGC 660
TGCAAACCTCT TCAGAACTGC CAAACAAAAC CCACTCACTG AGAAATCTGG TGTAACAAA 720
20 AAGGCTAAAA ACACAATTAT TCTTATTATT GTTGTGTTTG TTCTCTGTTT CACACCTTAC 780
CATGTTGCAA TTATTCAACA TATGATTAAG AAGCTTCGTT TCTCTAATTT CCTGGAATGT 840
AGCCAAAGAC ATTCGTTCCA GATTTCTCTG CACTTTACAG TATGCCTGAT GAACTTCAAT 900
TGCTGCATGG ACCCTTTTAT CTAATTCTTT GCATGTAAAG GGTATAAGAG AAAGGTTATG 960
AGGATGCTGA AACGGCAAGT CAGTGTATCG ATTTCTAGTG CTGTGAAGTC AGCCCCTGAA 1020
25 GAAAATTCAC GTGAAATGAC AGAAACGCAG ATGATGATAC ATTCCAAGTC TTCAAATGGA 1080
AAGTGA 1086

(207) INFORMATION FOR SEQ ID NO:206:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 361 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:206:

	Met	Asp	Ile	Gln	Met	Ala	Asn	Asn	Phe	Thr	Pro	Pro	Ser	Ala	Thr	Pro	
	1				5					10					15		
5	Gln	Gly	Asn	Asp	Cys	Asp	Leu	Tyr	Ala	His	His	Ser	Thr	Ala	Arg	Ile	
				20					25					30			
	Val	Met	Pro	Leu	His	Tyr	Ser	Leu	Val	Phe	Ile	Ile	Gly	Leu	Val	Gly	
				35				40					45				
10	Asn	Leu	Leu	Ala	Leu	Val	Val	Ile	Val	Gln	Asn	Arg	Lys	Lys	Ile	Asn	
		50					55					60					
	Ser	Thr	Thr	Leu	Tyr	Ser	Thr	Asn	Leu	Val	Ile	Ser	Asp	Ile	Leu	Phe	
	65					70					75				80		
	Thr	Thr	Ala	Leu	Pro	Thr	Arg	Ile	Ala	Tyr	Tyr	Ala	Met	Gly	Phe	Asp	
					85					90					95		
15	Trp	Arg	Ile	Gly	Asp	Ala	Leu	Cys	Arg	Ile	Thr	Ala	Leu	Val	Phe	Tyr	
				100					105					110			
	Ile	Asn	Thr	Tyr	Ala	Gly	Val	Asn	Phe	Met	Thr	Cys	Leu	Ser	Ile	Asp	
				115				120					125				
20	Arg	Phe	Ile	Ala	Val	Val	His	Pro	Leu	Arg	Tyr	Asn	Lys	Ile	Lys	Arg	
		130					135					140					
	Ile	Glu	His	Ala	Lys	Gly	Val	Cys	Ile	Phe	Val	Trp	Ile	Leu	Val	Phe	
	145					150					155				160		
	Ala	Gln	Thr	Leu	Pro	Leu	Leu	Ile	Asn	Pro	Met	Ser	Lys	Gln	Glu	Ala	
					165					170					175		
25	Glu	Arg	Ile	Thr	Cys	Met	Glu	Tyr	Pro	Asn	Phe	Glu	Glu	Thr	Lys	Ser	
				180					185					190			
	Leu	Pro	Trp	Ile	Leu	Leu	Gly	Ala	Cys	Phe	Ile	Gly	Tyr	Val	Leu	Pro	
				195				200					205				
30	Leu	Ile	Ile	Ile	Leu	Ile	Cys	Tyr	Ser	Gln	Ile	Cys	Cys	Lys	Leu	Phe	
		210					215					220					
	Arg	Thr	Ala	Lys	Gln	Asn	Pro	Leu	Thr	Glu	Lys	Ser	Gly	Val	Asn	Lys	
	225					230					235				240		
	Lys	Ala	Lys	Asn	Thr	Ile	Ile	Leu	Ile	Ile	Val	Val	Phe	Val	Leu	Cys	
				245						250					255		
35	Phe	Thr	Pro	Tyr	His	Val	Ala	Ile	Ile	Gln	His	Met	Ile	Lys	Lys	Leu	
				260						265					270		

163

Arg Phe Ser Asn Phe Leu Glu Cys Ser Gln Arg His Ser Phe Gln Ile
 275 280 285
 Ser Leu His Phe Thr Val Cys Leu Met Asn Phe Asn Cys Cys Met Asp
 290 295 300
 5 Pro Phe Ile Tyr Phe Phe Ala Cys Lys Gly Tyr Lys Arg Lys Val Met
 305 310 315 320
 Arg Met Leu Lys Arg Gln Val Ser Val Ser Ile Ser Ser Ala Val Lys
 325 330 335
 10 Ser Ala Pro Glu Glu Asn Ser Arg Glu Met Thr Glu Thr Gln Met Met
 340 345 350
 Ile His Ser Lys Ser Ser Asn Gly Lys
 355 360

(208) INFORMATION FOR SEQ ID NO:207:

15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1446 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:207:

ATGCGGTGGC TGTGGCCCTT GGCTGTCTCT CTTGCTGTGA TTTTGGCTGT GGGGCTAAGC 60
 AGGGTCTCTG GGGGTGCCCC CCTGCACCTG GGCAGGCACA GAGCCGAGAC CCAGGAGCAG 120
 CAGAGCCGAT CCAAGAGGGG CACCGAGGAT GAGGAGGCCA AGGGCGTGCA GCAGTATGTG 180
 CCTGAGGAGT GGGCGGAGTA CCCCCGGCCC ATTCACCCTG CTGGCCTGCA GCCAACCAAG 240
 25 CCCTTGGTGG CCACCAGCCC TAACCCCGAC AAGGATGGGG GCACCCGAGA CAGTGGGCAG 300
 GAACTGAGGG GCAATCTGAC AGGGGCACCA GGCAGAGGC TACAGATCCA GAACCCCTG 360
 TATCCGGTGA CCGAGAGCTC CTACAGTGCC TATGCCATCA TGCTTCTGGC GCTGGTGGTG 420
 TTTGCGGTGG GCATTGTGGG CAACCTGTCT GTCATGTGCA TCGTGTGGCA CAGCTACTAC 480
 CTGAAGAGCG CCTGGAAGTC CATCCTTGCC AGCCTGGCCC TCTGGGATTT TCTGGTCCTC 540
 30 TTTTCTTGCC TCCCTATTGT CATCTTCAAC GAGATCACCA AGCAGAGGCT ACTGGGTGAC 600
 GTTTCTTGTC GTGCCGTGCC CTTTCATGGG GTCTCCTCTC TGGGAGTCAC GACTTTCAGC 660
 CTCTGTGCCC TGGGCATTGA CCGCTTCCAC GTGGCCACCA GCACCCTGCC CAAGGTGAGG 720
 CCCATCGAGC GGTGCCAATC CATCCTGGCC AAGTTGGCTG TCATCTGGGT GGGCTCCATG 780

ACGCTGGCTG TGCCTGAGCT CCTGCTGTGG CAGCTGGCAC AGGAGCCTGC CCCACCATG 840
 GGCACCCTGG ACTCATGCAT CATGAAACCC TCAGCCAGCC TGCCCGAGTC CCTGTATTCA 900
 CTGGTGATGA CCTACCAGAA CGCCCGCATG TGGTGGTACT TTGGCTGCTA CTTCTGCCTG 960
 CCCATCCTCT TCACAGTCAC CTGCCAGCTG GTGACATGGC GGGTGCGAGG CCCTCCAGGG 1020
 5 AGGAAGTCAG AGTGCAGGGC CAGCAAGCAC GAGCAGTGTG AGAGCCAGCT CAAGAGCACC 1080
 GTGGTGGGCC TGACCGTGGT CTACGCCTTC TGCACCCTCC CAGAGAACGT CTGCAACATC 1140
 GTGGTGGCCT ACCTCTCCAC CGAGCTGACC CGCCAGACCC TGGACCTCCT GGGCCTCATC 1200
 AACCAGTTCT CCACCTTCTT CAAGGGCGCC ATCACCCAG TGCTGCTCCT TTGCATCTGC 1260
 AGGCCGCTGG GCCAGGCCTT CCTGGACTGC TGCTGCTGCT GCTGCTGTGA GGAGTGCGGC 1320
 10 GGGGCTTCGG AGGCCTCTGC TGCCAATGGG TCGGACAACA AGCTCAAGAC CGAGGTGTCC 1380
 TCTTCCATCT ACTTCCACAA GCCCAGGGAG TCACCCCCAC TCCTGCCCT GGGCACACCT 1440
 TGCTGA 1446

(209) INFORMATION FOR SEQ ID NO:208:

- 15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 481 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

- 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:208:

Met Arg Trp Leu Trp Pro Leu Ala Val Ser Leu Ala Val Ile Leu Ala
 1 5 10 15
 Val Gly Leu Ser Arg Val Ser Gly Gly Ala Pro Leu His Leu Gly Arg
 20 25 30
 25 His Arg Ala Glu Thr Gln Glu Gln Gln Ser Arg Ser Lys Arg Gly Thr
 35 40 45
 Glu Asp Glu Glu Ala Lys Gly Val Gln Gln Tyr Val Pro Glu Glu Trp
 50 55 60
 30 Ala Glu Tyr Pro Arg Pro Ile His Pro Ala Gly Leu Gln Pro Thr Lys
 65 70 75 80
 Pro Leu Val Ala Thr Ser Pro Asn Pro Asp Lys Asp Gly Gly Thr Pro
 85 90 95
 Asp Ser Gly Gln Glu Leu Arg Gly Asn Leu Thr Gly Ala Pro Gly Gln

165

				100					105					110		
	Arg	Leu	Gln	Ile	Gln	Asn	Pro	Leu	Tyr	Pro	Val	Thr	Glu	Ser	Ser	Tyr
			115						120				125			
5	Ser	Ala	Tyr	Ala	Ile	Met	Leu	Leu	Ala	Leu	Val	Val	Phe	Ala	Val	Gly
		130					135					140				
	Ile	Val	Gly	Asn	Leu	Ser	Val	Met	Cys	Ile	Val	Trp	His	Ser	Tyr	Tyr
	145					150					155					160
	Leu	Lys	Ser	Ala	Trp	Asn	Ser	Ile	Leu	Ala	Ser	Leu	Ala	Leu	Trp	Asp
					165					170					175	
10	Phe	Leu	Val	Leu	Phe	Phe	Cys	Leu	Pro	Ile	Val	Ile	Phe	Asn	Glu	Ile
				180					185						190	
	Thr	Lys	Gln	Arg	Leu	Leu	Gly	Asp	Val	Ser	Cys	Arg	Ala	Val	Pro	Phe
			195					200					205			
15	Met	Glu	Val	Ser	Ser	Leu	Gly	Val	Thr	Thr	Phe	Ser	Leu	Cys	Ala	Leu
		210					215					220				
	Gly	Ile	Asp	Arg	Phe	His	Val	Ala	Thr	Ser	Thr	Leu	Pro	Lys	Val	Arg
	225					230					235					240
	Pro	Ile	Glu	Arg	Cys	Gln	Ser	Ile	Leu	Ala	Lys	Leu	Ala	Val	Ile	Trp
					245					250					255	
20	Val	Gly	Ser	Met	Thr	Leu	Ala	Val	Pro	Glu	Leu	Leu	Leu	Trp	Gln	Leu
				260					265					270		
	Ala	Gln	Glu	Pro	Ala	Pro	Thr	Met	Gly	Thr	Leu	Asp	Ser	Cys	Ile	Met
			275					280					285			
25	Lys	Pro	Ser	Ala	Ser	Leu	Pro	Glu	Ser	Leu	Tyr	Ser	Leu	Val	Met	Thr
		290					295					300				
	Tyr	Gln	Asn	Ala	Arg	Met	Trp	Trp	Tyr	Phe	Gly	Cys	Tyr	Phe	Cys	Leu
	305					310					315					320
	Pro	Ile	Leu	Phe	Thr	Val	Thr	Cys	Gln	Leu	Val	Thr	Trp	Arg	Val	Arg
					325					330					335	
30	Gly	Pro	Pro	Gly	Arg	Lys	Ser	Glu	Cys	Arg	Ala	Ser	Lys	His	Glu	Gln
				340					345					350		
	Cys	Glu	Ser	Gln	Leu	Lys	Ser	Thr	Val	Val	Gly	Leu	Thr	Val	Val	Tyr
			355					360					365			
35	Ala	Phe	Cys	Thr	Leu	Pro	Glu	Asn	Val	Cys	Asn	Ile	Val	Val	Ala	Tyr
		370					375					380				
	Leu	Ser	Thr	Glu	Leu	Thr	Arg	Gln	Thr	Leu	Asp	Leu	Leu	Gly	Leu	Ile
	385					390					395					400

[illegible]

(210) INFORMATION FOR SEQ ID NO:209:

- 15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1101 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:209:

	ATGTGGAACG	CGACGCCCAG	CGAAGAGCCG	GGGTTCAACC	TCACACTGGC	CGACCTGGAC	60
	TGGGATGCTT	CCCCGGCAA	CGACTCGCTG	GGCGACGAGC	TGCTGCAGCT	CTTCCCCGCG	120
	CCGCTGCTGG	CGGGCGTCAC	AGCCACCTGC	GTGGCACTCT	TCGTGGTGGG	TATCGCTGGC	180
	AACCTGCTCA	CCATGCTGGT	GGTGTGCGCG	TTCCGCGAGC	TGCGCACCAC	CACCAACCTC	240
25	TACCTGTCCA	GCATGGCCTT	CTCCGATCTG	CTCATCTTCC	TCTGCATGCC	CCTGGACCTC	300
	GTTGCGCTCT	GGCAGTACCG	GCCCTGGAAC	TTCGGCGACC	TCCTCTGCAA	ACTCTTCCAA	360
	TTCGTCAGTG	AGAGCTGCAC	CTACGCCACG	GTGCTCACCA	TCACAGCGCT	GAGCGTCGAG	420
	CGCTACTTCG	CCATCTGCTT	CCCACTCCGG	GCCAAGGTGG	TGGTCACCAA	GGGGCGGGTG	480
	AAGCTGGTCA	TCTTCGTCAT	CTGGGCCGTG	GCCTTCTGCA	GCGCCGGGCC	CATCTTCGTG	540
30	CTAGTCGGGG	TGGAGCACGA	GAACGGCACC	GACCCTTGGG	ACACCAACGA	GTGCCGCCCC	600
	ACCGAGTTTG	CGGTGCGCTC	TGGACTGCTC	ACGGTCATGG	TGTGGGTGTC	CAGCATCTTC	660
	TTCTTCCTTC	CTGTCTTCTG	TCTCACGGTC	CTCTACAGTC	TCATCGGCAG	GAAGCTGTGG	720
	CGGAGGAGGC	GCGGCGATGC	TGTCGTGGGT	GCCTCGCTCA	GGGACCAGAA	CCACAAGCAA	780

167

ACCAAGAAAA TGCTGGCTGT AGTGGTGTTT GCCTTCATCC TCTGCTGGCT CCCCTTCCAC 840
 GTAGGGCGAT ATTTATTTTC CAAATCCTTT GAGCCTGGCT CCTTGGAGAT TGCTCAGATC 900
 AGCCAGTACT GCAACCTCGT GTCCTTTGTC CTCTTCTACC TCAGTGCTGC CATCAACCCC 960
 ATTCTGTACA ACATCATGTC CAAGAAGTAC CGGGTGGCAG TGTTTCAGACT TCTGGGATTC 1020
 5 GAACCCTTCT CCCAGAGAAA GCTCTCCACT CTGAAAGATG AAAGTTCTCG GGCCTGGACA 1080
 GAATCTAGTA TTAATACATG A 1101

(211) INFORMATION FOR SEQ ID NO:210:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 366 amino acids
 10 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:210:

15 Met Trp Asn Ala Thr Pro Ser Glu Glu Pro Gly Phe Asn Leu Thr Leu
 1 5 10 15
 Ala Asp Leu Asp Trp Asp Ala Ser Pro Gly Asn Asp Ser Leu Gly Asp
 20 25 30
 20 Glu Leu Leu Gln Leu Phe Pro Ala Pro Leu Leu Ala Gly Val Thr Ala
 35 40 45
 Thr Cys Val Ala Leu Phe Val Val Gly Ile Ala Gly Asn Leu Leu Thr
 50 55 60
 Met Leu Val Val Ser Arg Phe Arg Glu Leu Arg Thr Thr Thr Asn Leu
 65 70 75 80
 25 Tyr Leu Ser Ser Met Ala Phe Ser Asp Leu Leu Ile Phe Leu Cys Met
 85 90 95
 Pro Leu Asp Leu Val Arg Leu Trp Gln Tyr Arg Pro Trp Asn Phe Gly
 100 105 110
 30 Asp Leu Leu Cys Lys Leu Phe Gln Phe Val Ser Glu Ser Cys Thr Tyr
 115 120 125
 Ala Thr Val Leu Thr Ile Thr Ala Leu Ser Val Glu Arg Tyr Phe Ala
 130 135 140
 Ile Cys Phe Pro Leu Arg Ala Lys Val Val Val Thr Lys Gly Arg Val
 145 150 155 160
 35 Lys Leu Val Ile Phe Val Ile Trp Ala Val Ala Phe Cys Ser Ala Gly

165 170 175

Pro Ile Phe Val Leu Val Gly Val Glu His Glu Asn Gly Thr Asp Pro
180 185 190

5 Trp Asp Thr Asn Glu Cys Arg Pro Thr Glu Phe Ala Val Arg Ser Gly
195 200 205

Leu Leu Thr Val Met Val Trp Val Ser Ser Ile Phe Phe Phe Leu Pro
210 215 220

Val Phe Cys Leu Thr Val Leu Tyr Ser Leu Ile Gly Arg Lys Leu Trp
225 230 235 240

10 Arg Arg Arg Arg Gly Asp Ala Val Val Gly Ala Ser Leu Arg Asp Gln
245 250 255

Asn His Lys Gln Thr Lys Lys Met Leu Ala Val Val Val Phe Ala Phe
260 265 270

15 Ile Leu Cys Trp Leu Pro Phe His Val Gly Arg Tyr Leu Phe Ser Lys
275 280 285

Ser Phe Glu Pro Gly Ser Leu Glu Ile Ala Gln Ile Ser Gln Tyr Cys
290 295 300

Asn Leu Val Ser Phe Val Leu Phe Tyr Leu Ser Ala Ala Ile Asn Pro
305 310 315 320

20 Ile Leu Tyr Asn Ile Met Ser Lys Lys Tyr Arg Val Ala Val Phe Arg
325 330 335

Leu Leu Gly Phe Glu Pro Phe Ser Gln Arg Lys Leu Ser Thr Leu Lys
340 345 350

25 Asp Glu Ser Ser Arg Ala Trp Thr Glu Ser Ser Ile Asn Thr
355 360 365

(212) INFORMATION FOR SEQ ID NO:211:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1842 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:211:

ATGCGAGCCC CGGGCGCGCT TCTCGCCCGC ATGTCGCGGC TACTGCTTCT GCTACTGCTC 60

35 AAGGTGTCTG CCTCTTCTGC CCTCGGGGTC GCCCCTGCGT CCAGAAACGA AACTTGTCTG 120

GGGGAGAGCT GTGCACCTAC AGTGATCCAG CGCCGCGGCA GGGACGCCTG GGGACCGGGA 180

	AATTCTGCAA	GAGACGTTCT	GCGAGCCCGA	GCACCCAGGG	AGGAGCAGGG	GGCAGCGTTT	240
	CTTGCGGGAC	CCTCCTGGGA	CCTGCCGGCG	GCCCCGGGCC	GTGACCCGGC	TGCAGGCAGA	300
	GGGGCGGAGG	CGTCGGCAGC	CGGACCCCCG	GGACCTCCAA	CCAGGCCACC	TGGCCCCTGG	360
	AGGTGGAAG	GTGCTCGGGG	TCAGGAGCCT	TCTGAAACTT	TGGGGAGAGG	GAACCCACG	420
5	GCCCTCCAGC	TCTTCCTTCA	GATCTCAGAG	GAGGAAGAGA	AGGGTCCCAG	AGGCGCTGGC	480
	ATTTCCGGGC	GTAGCCAGGA	GCAGAGTGTG	AAGACAGTCC	CCGGAGCCAG	CGATCTTTTT	540
	TACTGGCCAA	GGAGAGCCGG	GAAACTCCAG	GGTTCCCACC	ACAAGCCCCT	GTCCAAGACG	600
	GCCAATGGAC	TGGCGGGGCA	CGAAGGGTGG	ACAATTGCAC	TCCCGGGCCG	GGCGCTGGCC	660
	CAGAATGGAT	CCTTGGGTGA	AGGAATCCAT	GAGCCTGGGG	GTCCCCGCCG	GGGAAACAGC	720
10	ACGAACCGGC	GTGTGAGACT	GAAGAACCCC	TTCTACCCGC	TGACCCAGGA	GTCCTATGGA	780
	GCCTACGCGG	TCATGTGTCT	GTCCGTGGTG	ATCTTCGGGA	CCGGCATCAT	TGGCAACCTG	840
	GCGGTGATGT	GCATCGTGTG	CCACAACACT	TACATGCGGA	GCATCTCCAA	CTCCCTCTTG	900
	GCCAACCTGG	CCTTCTGGGA	CTTTCTCATC	ATCTTCTTCT	GCCTTCCGCT	GGTCATCTTC	960
	CACGAGCTGA	CCAAGAAGTG	GCTGCTGGAG	GACTTCTCCT	GCAAGATCGT	GCCCTATATA	1020
15	GAGGTGCGCT	CTCTGGGAGT	CACCACTTTC	ACCTTATGTG	CTCTGTGCAT	AGACCGCTTC	1080
	CGTGCTGCCA	CCAACGTACA	GATGTACTAC	GAAATGATCG	AAAATTGTTC	CTCAACAAC	1140
	GCCAACTTG	CTGTTATATG	GGTGGGAGCT	CTATTGTTAG	CACTTCCAGA	AGTTGTTCTC	1200
	CGCCAGCTGA	GCAAGGAGGA	TTTGGGGTTT	AGTGGCCGAG	CTCCGGCAGA	AAGGTGCATT	1260
	ATTAAGATCT	CTCCTGATTT	ACCAGACACC	ATCTATGTTC	TAGCCCTCAC	CTACGACAGT	1320
20	GCGAGACTGT	GGTGGTATTT	TGGCTGTTAC	TTTTGTTTGC	CCACGCTTTT	CACCATCACC	1380
	TGCTCTCTAG	TGACTGCGAG	GAAAATCCGC	AAAGCAGAGA	AAGCCTGTAC	CCGAGGGAAT	1440
	AAACGGCAGA	TTCAACTAGA	GAGTCAGATG	AAGTGACAG	TAGTGGCACT	GACCATTTTA	1500
	TATGGATTTT	GCATTATTCC	TGAAAATATC	TGCAACATTG	TTACTGCCTA	CATGGCTACA	1560
	GGGGTTTCAC	AGCAGACAAT	GGACCTCCTT	AATATCATCA	GCCAGTTCCT	TTTGTTCTTT	1620
25	AAGTCCTGTG	TCACCCAGT	CCTCCTTTTC	TGTCTCTGCA	AACCCTTCAG	TCGGGCCTTC	1680
	ATGGAGTGCT	GCTGCTGTTG	CTGTGAGGAA	TGCATTGAGA	AGTCTTCAAC	GGTGACCAGT	1740
	GATGACAATG	ACAACGAGTA	CACCACGGAA	CTCGAACTCT	CGCCTTTCAG	TACCATACGC	1800
	CGTGAAATGT	CACTTTTGC	TTCTGTCGGA	ACTCATTGCT	GA		1842

(213) INFORMATION FOR SEQ ID NO:212:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 613 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:212:

10 Met Arg Ala Pro Gly Ala Leu Leu Ala Arg Met Ser Arg Leu Leu Leu
 1 5 10 15
 Leu Leu Leu Leu Lys Val Ser Ala Ser Ser Ala Leu Gly Val Ala Pro
 20 25 30
 Ala Ser Arg Asn Glu Thr Cys Leu Gly Glu Ser Cys Ala Pro Thr Val
 35 40 45
 15 Ile Gln Arg Arg Gly Arg Asp Ala Trp Gly Pro Gly Asn Ser Ala Arg
 50 55 60
 Asp Val Leu Arg Ala Arg Ala Pro Arg Glu Glu Gln Gly Ala Ala Phe
 65 70 75 80
 20 Leu Ala Gly Pro Ser Trp Asp Leu Pro Ala Ala Pro Gly Arg Asp Pro
 85 90 95
 Ala Ala Gly Arg Gly Ala Glu Ala Ser Ala Ala Gly Pro Pro Gly Pro
 100 105 110
 Pro Thr Arg Pro Pro Gly Pro Trp Arg Trp Lys Gly Ala Arg Gly Gln
 115 120 125
 25 Glu Pro Ser Glu Thr Leu Gly Arg Gly Asn Pro Thr Ala Leu Gln Leu
 130 135 140
 Phe Leu Gln Ile Ser Glu Glu Glu Glu Lys Gly Pro Arg Gly Ala Gly
 145 150 155 160
 30 Ile Ser Gly Arg Ser Gln Glu Gln Ser Val Lys Thr Val Pro Gly Ala
 165 170 175
 Ser Asp Leu Phe Tyr Trp Pro Arg Arg Ala Gly Lys Leu Gln Gly Ser
 180 185 190
 His His Lys Pro Leu Ser Lys Thr Ala Asn Gly Leu Ala Gly His Glu
 195 200 205
 35 Gly Trp Thr Ile Ala Leu Pro Gly Arg Ala Leu Ala Gln Asn Gly Ser
 210 215 220
 Leu Gly Glu Gly Ile His Glu Pro Gly Gly Pro Arg Arg Gly Asn Ser

171

	225		230		235		240									
	Thr	Asn	Arg	Arg	Val	Arg	Leu	Lys	Asn	Pro	Phe	Tyr	Pro	Leu	Thr	Gln
					245					250					255	
5	Glu	Ser	Tyr	Gly	Ala	Tyr	Ala	Val	Met	Cys	Leu	Ser	Val	Val	Ile	Phe
				260					265					270		
	Gly	Thr	Gly	Ile	Ile	Gly	Asn	Leu	Ala	Val	Met	Cys	Ile	Val	Cys	His
				275				280						285		
	Asn	Tyr	Tyr	Met	Arg	Ser	Ile	Ser	Asn	Ser	Leu	Leu	Ala	Asn	Leu	Ala
		290					295					300				
10	Phe	Trp	Asp	Phe	Leu	Ile	Ile	Phe	Phe	Cys	Leu	Pro	Leu	Val	Ile	Phe
	305					310					315				320	
	His	Glu	Leu	Thr	Lys	Lys	Trp	Leu	Leu	Glu	Asp	Phe	Ser	Cys	Lys	Ile
					325					330					335	
15	Val	Pro	Tyr	Ile	Glu	Val	Ala	Ser	Leu	Gly	Val	Thr	Thr	Phe	Thr	Leu
				340					345					350		
	Cys	Ala	Leu	Cys	Ile	Asp	Arg	Phe	Arg	Ala	Ala	Thr	Asn	Val	Gln	Met
		355						360					365			
	Tyr	Tyr	Glu	Met	Ile	Glu	Asn	Cys	Ser	Ser	Thr	Thr	Ala	Lys	Leu	Ala
		370					375					380				
20	Val	Ile	Trp	Val	Gly	Ala	Leu	Leu	Leu	Ala	Leu	Pro	Glu	Val	Val	Leu
	385					390					395				400	
	Arg	Gln	Leu	Ser	Lys	Glu	Asp	Leu	Gly	Phe	Ser	Gly	Arg	Ala	Pro	Ala
					405					410					415	
25	Glu	Arg	Cys	Ile	Ile	Lys	Ile	Ser	Pro	Asp	Leu	Pro	Asp	Thr	Ile	Tyr
				420					425					430		
	Val	Leu	Ala	Leu	Thr	Tyr	Asp	Ser	Ala	Arg	Leu	Trp	Trp	Tyr	Phe	Gly
		435						440					445			
	Cys	Tyr	Phe	Cys	Leu	Pro	Thr	Leu	Phe	Thr	Ile	Thr	Cys	Ser	Leu	Val
		450					455					460				
30	Thr	Ala	Arg	Lys	Ile	Arg	Lys	Ala	Glu	Lys	Ala	Cys	Thr	Arg	Gly	Asn
	465					470					475				480	
	Lys	Arg	Gln	Ile	Gln	Leu	Glu	Ser	Gln	Met	Lys	Cys	Thr	Val	Val	Ala
				485					490					495		
35	Leu	Thr	Ile	Leu	Tyr	Gly	Phe	Cys	Ile	Ile	Pro	Glu	Asn	Ile	Cys	Asn
				500					505					510		
	Ile	Val	Thr	Ala	Tyr	Met	Ala	Thr	Gly	Val	Ser	Gln	Gln	Thr	Met	Asp
		515						520					525			

Leu Leu Asn Ile Ile Ser Gln Phe Leu Leu Phe Phe Lys Ser Cys Val
 530 535 540
 Thr Pro Val Leu Leu Phe Cys Leu Cys Lys Pro Phe Ser Arg Ala Phe
 545 550 555 560
 5 Met Glu Cys Cys Cys Cys Cys Glu Glu Cys Ile Gln Lys Ser Ser
 565 570 575
 Thr Val Thr Ser Asp Asp Asn Asp Asn Glu Tyr Thr Thr Glu Leu Glu
 580 585 590
 10 Leu Ser Pro Phe Ser Thr Ile Arg Arg Glu Met Ser Thr Phe Ala Ser
 595 600 605
 Val Gly Thr His Cys
 610

(214) INFORMATION FOR SEQ ID NO:213:

15 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1248 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:213:

ATGGTTTTTG CTCACAGAAT GGATAACAGC AAGCCACATT TGATTATTCC TAACTTCTG 60
 GTGCCCCTCC AAAACCGCAG CTGCACTGAA ACAGCCACAC CTCTGCCAAG CCAATACCTG 120
 ATGGAATTAA GTGAGGAGCA CAGTTGGATG AGCAACCAAA CAGACCTTCA CTATGTGCTG 180
 AAACCCGGGG AAGTGGCCAC AGCCAGCATC TTCTTTGGGA TTCTGTGGTT GTTTTCTATC 240
 25 TTCGGCAATT CCCTGGTTTG TTTGGTCATC CATAGGAGTA GGAGGACTCA GTCTACCACC 300
 AACTACTTTG TGGTCTCCAT GGCATGTGCT GACCTTCTCA TCAGCGTTGC CAGCACGCCT 360
 TTCGTCCTGC TCCAGTTCAC CACTGGAAGG TGGACGCTGG GTAGTGCAAC GTGCAAGGTT 420
 GTGCGATATT TTCAATATCT CACTCCAGGT GTCCAGATCT ACGTTCTCCT CTCCATCTGC 480
 ATAGACCGGT TCTACACCAT CGTCTATCCT CTGAGCTTCA AGGTGTCCAG AGAAAAAGCC 540
 30 AAGAAAATGA TTGCGGCATC GTGGATCTTT GATGCAGGCT TTGTGACCCC TGTGCTCTTT 600
 TTCTATGGCT CCAACTGGGA CAGTCATTGT AACTATTTCC TCCCCTCCTC TTGGGAAGGC 660
 ACTGCCTACA CTGTCATCCA CTTCTTGGTG GGCTTTGTGA TTCCATCTGT CCTCATAATT 720
 TTATTTTACC AAAAGGTCAT AAAATATATT TGGAGAATAG GCACAGATGG CCGAACGGTG 780

173

AGGAGGACAA TGAACATTGT CCCTCGGACA AAAGTGAAAA CTAAAAAGAT GTTCCTCATT 840
 TTAAATCTGT TGTTTTTGCT CTCCTGGCTG CCTTTTCATG TAGCTCAGCT ATGGCACCCC 900
 CATGAACAAG ACTATAAGAA AAGTTCCCTT GTTTTCACAG CTATCACATG GATATCCTTT 960
 AGTTCTTCAG CCTCTAAACC TACTCTGTAT TCAATTTATA ATGCCAATTT TCGGAGAGGG 1020
 5 ATGAAAGAGA CTTTTTGCAT GTCCTCTATG AAATGTTACC GAAGCAATGC CTATACTATC 1080
 ACAACAAGTT CAAGGATGGC CAAAAAAAC TACGTTGGCA TTTCAGAAAT CCCTTCCATG 1140
 GCCAAACTA TTACCAAAGA CTCGATCTAT GACTCATTTG ACAGAGAAGC CAAGGAAAAA 1200
 AAGCTTGCTT GGCCCATTA CTCAAATCCA CCAAATACTT TTGTCTAA 1248

(215) INFORMATION FOR SEQ ID NO:214:

- 10 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 415 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant

- 15 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:214:

	Met	Val	Phe	Ala	His	Arg	Met	Asp	Asn	Ser	Lys	Pro	His	Leu	Ile	Ile	
	1				5					10					15		
	Pro	Thr	Leu	Leu	Val	Pro	Leu	Gln	Asn	Arg	Ser	Cys	Thr	Glu	Thr	Ala	
20				20					25					30			
	Thr	Pro	Leu	Pro	Ser	Gln	Tyr	Leu	Met	Glu	Leu	Ser	Glu	Glu	His	Ser	
				35				40					45				
	Trp	Met	Ser	Asn	Gln	Thr	Asp	Leu	His	Tyr	Val	Leu	Lys	Pro	Gly	Glu	
		50				55					60						
25	Val	Ala	Thr	Ala	Ser	Ile	Phe	Phe	Gly	Ile	Leu	Trp	Leu	Phe	Ser	Ile	
	65					70				75					80		
	Phe	Gly	Asn	Ser	Leu	Val	Cys	Leu	Val	Ile	His	Arg	Ser	Arg	Arg	Thr	
					85					90					95		
	Gln	Ser	Thr	Thr	Asn	Tyr	Phe	Val	Val	Ser	Met	Ala	Cys	Ala	Asp	Leu	
30				100				105					110				
	Leu	Ile	Ser	Val	Ala	Ser	Thr	Pro	Phe	Val	Leu	Leu	Gln	Phe	Thr	Thr	
		115					120					125					
	Gly	Arg	Trp	Thr	Leu	Gly	Ser	Ala	Thr	Cys	Lys	Val	Val	Arg	Tyr	Phe	
		130				135						140					

	Gln Tyr Leu Thr Pro Gly Val Gln Ile Tyr Val Leu Leu Ser Ile Cys	
	145	150 155 160
	Ile Asp Arg Phe Tyr Thr Ile Val Tyr Pro Leu Ser Phe Lys Val Ser	
		165 170 175
5	Arg Glu Lys Ala Lys Lys Met Ile Ala Ala Ser Trp Ile Phe Asp Ala	
		180 185 190
	Gly Phe Val Thr Pro Val Leu Phe Phe Tyr Gly Ser Asn Trp Asp Ser	
		195 200 205
10	His Cys Asn Tyr Phe Leu Pro Ser Ser Trp Glu Gly Thr Ala Tyr Thr	
		210 215 220
	Val Ile His Phe Leu Val Gly Phe Val Ile Pro Ser Val Leu Ile Ile	
		225 230 235 240
	Leu Phe Tyr Gln Lys Val Ile Lys Tyr Ile Trp Arg Ile Gly Thr Asp	
		245 250 255
15	Gly Arg Thr Val Arg Arg Thr Met Asn Ile Val Pro Arg Thr Lys Val	
		260 265 270
	Lys Thr Lys Lys Met Phe Leu Ile Leu Asn Leu Leu Phe Leu Leu Ser	
		275 280 285
20	Trp Leu Pro Phe His Val Ala Gln Leu Trp His Pro His Glu Gln Asp	
		290 295 300
	Tyr Lys Lys Ser Ser Leu Val Phe Thr Ala Ile Thr Trp Ile Ser Phe	
		305 310 315 320
	Ser Ser Ser Ala Ser Lys Pro Thr Leu Tyr Ser Ile Tyr Asn Ala Asn	
		325 330 335
25	Phe Arg Arg Gly Met Lys Glu Thr Phe Cys Met Ser Ser Met Lys Cys	
		340 345 350
	Tyr Arg Ser Asn Ala Tyr Thr Ile Thr Thr Ser Ser Arg Met Ala Lys	
		355 360 365
30	Lys Asn Tyr Val Gly Ile Ser Glu Ile Pro Ser Met Ala Lys Thr Ile	
		370 375 380
	Thr Lys Asp Ser Ile Tyr Asp Ser Phe Asp Arg Glu Ala Lys Glu Lys	
		385 390 395 400
	Lys Leu Ala Trp Pro Ile Asn Ser Asn Pro Pro Asn Thr Phe Val	
		405 410 415

35 (216) INFORMATION FOR SEQ ID NO:215:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1842 base pairs

175

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:215:

	ATGGGGCCCA CCCTAGCGGT TCCCACCCCC TATGGCTGTA TTGGCTGTAA GCTACCCCAG	60
	CCAGAATACC CACCGGCTCT AATCATCTTT ATGTTCTGCG CGATGGTTAT CACCATCGTT	120
	GTAGACCTAA TCGGCAACTC CATGGTCATT TTGGCTGTGA CGAAGAACAA GAAGCTCCGG	180
	AATTCTGGCA ACATCTTCGT GGTCAGTCTC TCTGTGGCCG ATATGCTGGT GGCCATCTAC	240
10	CCATACCCTT TGATGCTGCA TGCCATGTCC ATTGGGGGCT GGGATCTGAG CCAGTTACAG	300
	TGCCAGATGG TCGGGTTCAT CACAGGGCTG AGTGTGGTCG GCTCCATCTT CAACATCGTG	360
	GCAATCGCTA TCAACCGTTA CTGCTACATC TGCCACAGCC TCCAGTACGA ACGGATCTTC	420
	AGTGTGCGCA ATACCTGCAT CTACCTGGTC ATCACCTGGA TCATGACCGT CCTGGCTGTC	480
	CTGCCCAACA TGTACATTGG CACCATCGAG TACGATCCTC GCACCTACAC CTGCATCTTC	540
15	AACTATCTGA ACAACCCTGT CTTCACTGTT ACCATCGTCT GCATCCACTT CGTCTCCCT	600
	CTCTCATCG TGGGTTTCTG CTACGTGAGG ATCTGGACCA AAGTGCTGGC GGCCCGTGAC	660
	CCTGCAGGGC AGAATCCTGA CAACCAACTT GCTGAGGTTT GCAATAAACT AACCATGTTT	720
	GTGATCTTCC TCCTCTTTC AGTGTGCTGG TGCCCTATCA ACGTGCTCAC TGTCTTGGTG	780
	GCTGTCAGTC CGAAGGAGAT GGCAGGCAAG ATCCCCAACT GGCTTTATCT TGCAGCCTAC	840
20	TTCATAGCCT ACTTCAACAG CTGCCTCAAC GCTGTGATCT ACGGGCTCCT CAATGAGAAT	900
	TTCCGAAGAG AATACTGGAC CATCTTCCAT GCTATGCGGC ACCCTATCAT ATTCTTCTCT	960
	GGCCTCATCA GTGATATTCG TGAGATGCAG GAGGCCCCGTA CCCTGGCCCCG CGCCCGTGCC	1020
	CATGCTCGCG ACCAAGCTCG TGAACAAGAC CGTGCCCATG CCTGTCCTGC TGTGGAGGAA	1080
	ACCCCGATGA ATGTCCGGAA TGTTCCATTA CCTGGTGATG CTGCAGCTGG CCACCCCGAC	1140
25	CGTGCCCTCTG GCCACCCTAA GCCCCATTCC AGATCCTCCT CTGCCTATCG CAAATCTGCC	1200
	TCTACCCACC ACAAGTCTGT CTTTAGCCAC TCCAAGGCTG CCTCTGGTCA CCTCAAGCCT	1260
	GTCTCTGGCC ACTCCAAGCC TGCCTCTGGT CACCCCAAGT CTGCCACTGT CTACCCTAAG	1320
	CCTGCCTCTG TCCATTTCAA GGCTGACTCT GTCCATTTCA AGGGTGACTC TGTCCATTTT	1380
	AAGCCTGACT CTGTTTATTT CAAGCCTGCT TCCAGCAACC CCAAGCCCAT CACTGGCCAC	1440

CATGTCTCTG CTGGCAGCCA CTCCAAGTCT GCCTTCAATG CTGCCACCAG CCACCCTAAA 1500
 CCCATCAAGC CAGCTACCAG CCATGCTGAG CCCACCACTG CTGACTATCC CAAGCCTGCC 1560
 ACTACCAGCC ACCCTAAGCC CGCTGCTGCT GACAACCCTG AGCTCTCTGC CTCCCATTGC 1620
 CCCGAGATCC CTGCCATTGC CCACCCTGTG TCTGACGACA GTGACCTCCC TGAGTCGGCC 1680
 5 TCTAGCCCTG CCGCTGGGCC CACCAAGCCT GCTGCCAGCC AGCTGGAGTC TGACACCATC 1740
 GCTGACCTTC CTGACCCTAC TGTAGTCACT ACCAGTACCA ATGATTACCA TGATGTCGTG 1800
 GTTGTGATG TTGAAGATGA TCCTGATGAA ATGGCTGTGT GA 1842

(217) INFORMATION FOR SEQ ID NO:216:

10 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 613 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

 (ii) MOLECULE TYPE: protein

 15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:216:

 Met Gly Pro Thr Leu Ala Val Pro Thr Pro Tyr Gly Cys Ile Gly Cys
 1 5 10 15
 Lys Leu Pro Gln Pro Glu Tyr Pro Pro Ala Leu Ile Ile Phe Met Phe
 20 25 30
 20 Cys Ala Met Val Ile Thr Ile Val Val Asp Leu Ile Gly Asn Ser Met
 35 40 45
 Val Ile Leu Ala Val Thr Lys Asn Lys Lys Leu Arg Asn Ser Gly Asn
 50 55 60
 25 Ile Phe Val Val Ser Leu Ser Val Ala Asp Met Leu Val Ala Ile Tyr
 65 70 75 80
 Pro Tyr Pro Leu Met Leu His Ala Met Ser Ile Gly Gly Trp Asp Leu
 85 90 95
 Ser Gln Leu Gln Cys Gln Met Val Gly Phe Ile Thr Gly Leu Ser Val
 100 105 110
 30 Val Gly Ser Ile Phe Asn Ile Val Ala Ile Ala Ile Asn Arg Tyr Cys
 115 120 125
 Tyr Ile Cys His Ser Leu Gln Tyr Glu Arg Ile Phe Ser Val Arg Asn
 130 135 140
 35 Thr Cys Ile Tyr Leu Val Ile Thr Trp Ile Met Thr Val Leu Ala Val
 145 150 155 160

177

	Leu	Pro	Asn	Met	Tyr	Ile	Gly	Thr	Ile	Glu	Tyr	Asp	Pro	Arg	Thr	Tyr	
					165					170					175		
	Thr	Cys	Ile	Phe	Asn	Tyr	Leu	Asn	Asn	Pro	Val	Phe	Thr	Val	Thr	Ile	
				180					185					190			
5	Val	Cys	Ile	His	Phe	Val	Leu	Pro	Leu	Leu	Ile	Val	Gly	Phe	Cys	Tyr	
			195					200					205				
	Val	Arg	Ile	Trp	Thr	Lys	Val	Leu	Ala	Ala	Arg	Asp	Pro	Ala	Gly	Gln	
		210					215					220					
10	Asn	Pro	Asp	Asn	Gln	Leu	Ala	Glu	Val	Arg	Asn	Lys	Leu	Thr	Met	Phe	
		225				230					235					240	
	Val	Ile	Phe	Leu	Leu	Phe	Ala	Val	Cys	Trp	Cys	Pro	Ile	Asn	Val	Leu	
				245						250					255		
	Thr	Val	Leu	Val	Ala	Val	Ser	Pro	Lys	Glu	Met	Ala	Gly	Lys	Ile	Pro	
			260						265					270			
15	Asn	Trp	Leu	Tyr	Leu	Ala	Ala	Tyr	Phe	Ile	Ala	Tyr	Phe	Asn	Ser	Cys	
			275					280					285				
	Leu	Asn	Ala	Val	Ile	Tyr	Gly	Leu	Leu	Asn	Glu	Asn	Phe	Arg	Arg	Glu	
		290					295					300					
20	Tyr	Trp	Thr	Ile	Phe	His	Ala	Met	Arg	His	Pro	Ile	Ile	Phe	Phe	Ser	
		305				310					315					320	
	Gly	Leu	Ile	Ser	Asp	Ile	Arg	Glu	Met	Gln	Glu	Ala	Arg	Thr	Leu	Ala	
				325						330					335		
	Arg	Ala	Arg	Ala	His	Ala	Arg	Asp	Gln	Ala	Arg	Glu	Gln	Asp	Arg	Ala	
				340					345					350			
25	His	Ala	Cys	Pro	Ala	Val	Glu	Glu	Thr	Pro	Met	Asn	Val	Arg	Asn	Val	
			355					360					365				
	Pro	Leu	Pro	Gly	Asp	Ala	Ala	Ala	Gly	His	Pro	Asp	Arg	Ala	Ser	Gly	
		370					375					380					
30	His	Pro	Lys	Pro	His	Ser	Arg	Ser	Ser	Ser	Ala	Tyr	Arg	Lys	Ser	Ala	
		385				390					395					400	
	Ser	Thr	His	His	Lys	Ser	Val	Phe	Ser	His	Ser	Lys	Ala	Ala	Ser	Gly	
				405						410					415		
	His	Leu	Lys	Pro	Val	Ser	Gly	His	Ser	Lys	Pro	Ala	Ser	Gly	His	Pro	
				420					425					430			
35	Lys	Ser	Ala	Thr	Val	Tyr	Pro	Lys	Pro	Ala	Ser	Val	His	Phe	Lys	Ala	
			435					440					445				
	Asp	Ser	Val	His	Phe	Lys	Gly	Asp	Ser	Val	His	Phe	Lys	Pro	Asp	Ser	

450 455 460
 Val His Phe Lys Pro Ala Ser Ser Asn Pro Lys Pro Ile Thr Gly His
 465 470 475 480
 5 His Val Ser Ala Gly Ser His Ser Lys Ser Ala Phe Asn Ala Ala Thr
 485 490 495
 Ser His Pro Lys Pro Ile Lys Pro Ala Thr Ser His Ala Glu Pro Thr
 500 505 510
 Thr Ala Asp Tyr Pro Lys Pro Ala Thr Thr Ser His Pro Lys Pro Ala
 515 520 525
 10 Ala Ala Asp Asn Pro Glu Leu Ser Ala Ser His Cys Pro Glu Ile Pro
 530 535 540
 Ala Ile Ala His Pro Val Ser Asp Asp Ser Asp Leu Pro Glu Ser Ala
 545 550 555 560
 15 Ser Ser Pro Ala Ala Gly Pro Thr Lys Pro Ala Ala Ser Gln Leu Glu
 565 570 575
 Ser Asp Thr Ile Ala Asp Leu Pro Asp Pro Thr Val Val Thr Thr Ser
 580 585 590
 Thr Asn Asp Tyr His Asp Val Val Val Val Asp Val Glu Asp Asp Pro
 595 600 605
 20 Asp Glu Met Ala Val
 610

(218) INFORMATION FOR SEQ ID NO:217:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1854 base pairs
 25 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:217:

30 ATGGGGCCCA CCCTAGCGGT TCCCACCCCC TATGGCTGTA TTGGCTGTAA GCTACCCCAG 60
 CCAGAATACC CACCGGCTCT AATCATCTTT ATGTTCTGCG CGATGGTTAT CACCATCGTT 120
 GTAGACCTAA TCGGCAACTC CATGGTCATT TTGGCTGTGA CGAAGAACAA GAAGCTCCGG 180
 AATTCTGGCA ACATCTTCGT GGTCAGTCTC TCTGTGGCCG ATATGCTGGT GGCCATCTAC 240
 CCATACCCCTT TGATGCTGCA TGCCATGTCC ATTGGGGGCT GGGATCTGAG CCAGTTACAG 300
 35 TGCCAGATGG TCGGGTTCAT CACAGGGCTG AGTGTGGTCG GCTCCATCTT CAACATCGTG 360

	GCAATCGCTA TCAACCGTTA CTGCTACATC TGCCACAGCC TCCAGTACGA ACGGATCTTC	420
	AGTGTGCGCA ATACCTGCAT CTACCTGGTC ATCACCTGGA TCATGACCGT CCTGGCTGTC	480
	CTGCCCCAACA TGTACATTGG CACCATCGAG TACGATCCTC GCACCTACAC CTGCATCTTC	540
	AACTATCTGA ACAACCCTGT CTTCACTGTT ACCATCGTCT GCATCCACTT CGTCCTCCCT	600
5	CTCCTCATCG TGGGTTTCTG CTACGTGAGG ATCTGGACCA AAGTGCTGGC GGCCCGTGAC	660
	CCTGCAGGGC AGAATCCTGA CAACCAACTT GCTGAGGTTC GCAATAAACT AACCATGTTT	720
	GTGATCTTCC TCCTCTTTGC AGTGTGCTGG TGCCCTATCA ACGTGCTCAC TGTCTTGGTG	780
	GCTGTCAGTC CGAAGGAGAT GGCAGGCAAG ATCCCCAACT GGCTTTATCT TGCAGCCTAC	840
	TTCATAGCCT ACTTCAACAG CTGCCTCAAC GCTGTGATCT ACGGGCTCCT CAATGAGAAT	900
10	TTCCGAAGAG AATACTGGAC CATCTTCCAT GCTATGCGGC ACCCTATCAT ATTCTTCTCT	960
	GGCCTCATCA GTGATATTCG TGAGATGCAG GAGGCCCGTA CCCTGGCCCG CGCCCGTGCC	1020
	CATGCTCGCG ACCAAGCTCG TGAACAAGAC CGTGCCCATG CCTGTCCTGC TGTGGAGGAA	1080
	ACCCCGATGA ATGTCCGGAA TGTTCCATTA CCTGGTGATG CTGCAGCTGG CCACCCGAC	1140
	CGTGCCTCTG GCCACCCTAA GCCCCATTCC AGATCCTCCT CTGCCTATCG CAAATCTGCC	1200
15	TCTACCCACC ACAAGTCTGT CTTTAGCCAC TCCAAGGCTG CCTCTGGTCA CCTCAAGCCT	1260
	GTCTCTGGCC ACTCCAAGCC TGCCTCTGGT CACCCCAAGT CTGCCACTGT CTACCCTAAG	1320
	CCTGCCTCTG TCCATTTCAA GGCTGACTCT GTCCATTTCA AGGGTGACTC TGTCCATTTT	1380
	AAGCCTGACT CTGTTTCAAT CAAGCCTGCT TCCAGCAACC CCAAGCCCAT CACTGGCCAC	1440
	CATGTCTCTG CTGGCAGCCA CTCCAAGTCT GCCTTCAGTG CTGCCACCAG CCACCCTAAA	1500
20	CCCACCACTG GCCACATCAA GCCAGCTACC AGCCATGCTG AGCCCACCAC TGCTGACTAT	1560
	CCCAAGCCTG CCACTACCAG CCACCCTAAG CCCACTGCTG CTGACAACCC TGAGCTCTCT	1620
	GCCTCCCATT GCCCCGAGAT CCCTGCCATT GCCCACCCTG TGTCTGACGA CAGTGACCTC	1680
	CCTGAGTCGG CCTCTAGCCC TGCCGCTGGG CCCACCAAGC CTGCTGCCAG CCAGCTGGAG	1740
	TCTGACACCA TCGCTGACCT TCCTGACCCT ACTGTAGTCA CTACCAGTAC CAATGATTAC	1800
25	CATGATGTCG TGGTTGTTGA TGTGAAGAT GATCCTGATG AAATGGCTGT GTGA	1854

(219) INFORMATION FOR SEQ ID NO:218:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 617 amino acids
 - (B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:218:

5	Met Gly Pro Thr Leu Ala Val Pro Thr Pro Tyr Gly Cys Ile Gly Cys	1	5	10	15
	Lys Leu Pro Gln Pro Glu Tyr Pro Pro Ala Leu Ile Ile Phe Met Phe	20	25	30	
10	Cys Ala Met Val Ile Thr Ile Val Val Asp Leu Ile Gly Asn Ser Met	35	40	45	
	Val Ile Leu Ala Val Thr Lys Asn Lys Lys Leu Arg Asn Ser Gly Asn	50	55	60	
	Ile Phe Val Val Ser Leu Ser Val Ala Asp Met Leu Val Ala Ile Tyr	65	70	75	80
15	Pro Tyr Pro Leu Met Leu His Ala Met Ser Ile Gly Gly Trp Asp Leu	85	90	95	
	Ser Gln Leu Gln Cys Gln Met Val Gly Phe Ile Thr Gly Leu Ser Val	100	105	110	
20	Val Gly Ser Ile Phe Asn Ile Val Ala Ile Ala Ile Asn Arg Tyr Cys	115	120	125	
	Tyr Ile Cys His Ser Leu Gln Tyr Glu Arg Ile Phe Ser Val Arg Asn	130	135	140	
	Thr Cys Ile Tyr Leu Val Ile Thr Trp Ile Met Thr Val Leu Ala Val	145	150	155	160
25	Leu Pro Asn Met Tyr Ile Gly Thr Ile Glu Tyr Asp Pro Arg Thr Tyr	165	170	175	
	Thr Cys Ile Phe Asn Tyr Leu Asn Asn Pro Val Phe Thr Val Thr Ile	180	185	190	
30	Val Cys Ile His Phe Val Leu Pro Leu Leu Ile Val Gly Phe Cys Tyr	195	200	205	
	Val Arg Ile Trp Thr Lys Val Leu Ala Ala Arg Asp Pro Ala Gly Gln	210	215	220	
	Asn Pro Asp Asn Gln Leu Ala Glu Val Arg Asn Lys Leu Thr Met Phe	225	230	235	240
35	Val Ile Phe Leu Leu Phe Ala Val Cys Trp Cys Pro Ile Asn Val Leu	245	250	255	

181

	Thr	Val	Leu	Val	Ala	Val	Ser	Pro	Lys	Glu	Met	Ala	Gly	Lys	Ile	Pro	
				260					265					270			
	Asn	Trp	Leu	Tyr	Leu	Ala	Ala	Tyr	Phe	Ile	Ala	Tyr	Phe	Asn	Ser	Cys	
			275					280					285				
5	Leu	Asn	Ala	Val	Ile	Tyr	Gly	Leu	Leu	Asn	Glu	Asn	Phe	Arg	Arg	Glu	
		290					295					300					
	Tyr	Trp	Thr	Ile	Phe	His	Ala	Met	Arg	His	Pro	Ile	Ile	Phe	Phe	Ser	
	305					310					315					320	
10	Gly	Leu	Ile	Ser	Asp	Ile	Arg	Glu	Met	Gln	Glu	Ala	Arg	Thr	Leu	Ala	
					325					330					335		
	Arg	Ala	Arg	Ala	His	Ala	Arg	Asp	Gln	Ala	Arg	Glu	Gln	Asp	Arg	Ala	
				340					345					350			
	His	Ala	Cys	Pro	Ala	Val	Glu	Glu	Thr	Pro	Met	Asn	Val	Arg	Asn	Val	
			355					360					365				
15	Pro	Leu	Pro	Gly	Asp	Ala	Ala	Ala	Gly	His	Pro	Asp	Arg	Ala	Ser	Gly	
		370				375						380					
	His	Pro	Lys	Pro	His	Ser	Arg	Ser	Ser	Ser	Ala	Tyr	Arg	Lys	Ser	Ala	
	385					390					395					400	
20	Ser	Thr	His	His	Lys	Ser	Val	Phe	Ser	His	Ser	Lys	Ala	Ala	Ser	Gly	
					405					410					415		
	His	Leu	Lys	Pro	Val	Ser	Gly	His	Ser	Lys	Pro	Ala	Ser	Gly	His	Pro	
				420					425					430			
	Lys	Ser	Ala	Thr	Val	Tyr	Pro	Lys	Pro	Ala	Ser	Val	His	Phe	Lys	Ala	
			435					440					445				
25	Asp	Ser	Val	His	Phe	Lys	Gly	Asp	Ser	Val	His	Phe	Lys	Pro	Asp	Ser	
		450					455					460					
	Val	His	Phe	Lys	Pro	Ala	Ser	Ser	Asn	Pro	Lys	Pro	Ile	Thr	Gly	His	
	465					470					475					480	
30	His	Val	Ser	Ala	Gly	Ser	His	Ser	Lys	Ser	Ala	Phe	Ser	Ala	Ala	Thr	
					485					490					495		
	Ser	His	Pro	Lys	Pro	Thr	Thr	Gly	His	Ile	Lys	Pro	Ala	Thr	Ser	His	
				500					505					510			
	Ala	Glu	Pro	Thr	Thr	Ala	Asp	Tyr	Pro	Lys	Pro	Ala	Thr	Thr	Ser	His	
			515					520					525				
35	Pro	Lys	Pro	Thr	Ala	Ala	Asp	Asn	Pro	Glu	Leu	Ser	Ala	Ser	His	Cys	
		530					535					540					
	Pro	Glu	Ile	Pro	Ala	Ile	Ala	His	Pro	Val	Ser	Asp	Asp	Ser	Asp	Leu	

545 550 555 560
 Pro Glu Ser Ala Ser Ser Pro Ala Ala Gly Pro Thr Lys Pro Ala Ala
 565 570 575
 Ser Gln Leu Glu Ser Asp Thr Ile Ala Asp Leu Pro Asp Pro Thr Val
 5 580 585 590
 Val Thr Thr Ser Thr Asn Asp Tyr His Asp Val Val Val Val Asp Val
 595 600 605
 Glu Asp Asp Pro Asp Glu Met Ala Val
 610 615

10 (220) INFORMATION FOR SEQ ID NO:219:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1548 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

15 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:219:

ATGGGACATA ACGGGAGCTG GATCTCTCCA AATGCCAGCG AGCCGCACAA CGCGTCCGGC 60
 GCCGAGGCTG CGGGTGTGAA CCGCAGCGCG CTCGGGGAGT TCGGCGAGGC GCAGCTGTAC 120
 20 CGCCAGTTCA CCACCACCGT GCAGGTCGTC ATCTTCATAG GCTCGCTGCT CGGAAACTTC 180
 ATGGTGTAT GGTCAACTTG CCGCACAACC GTGTTCAAAT CTGTCACCAA CAGGTTCAAT 240
 AAAAACCTGG CCTGCTCGGG GATTTGTGCC AGCCTGGTCT GTGTGCCCTT CGACATCATC 300
 CTCAGCACCA GTCCTCACTG TTGCTGGTGG ATCTACACCA TGCTCTTCTG CAAGGTCGTC 360
 AAATTTTTCG ACAAAGTATT CTGCTCTGTG ACCATCCTCA GCTTCCCTGC TATTGCTTTG 420
 25 GACAGGTACT ACTCAGTCCT CTATCCACTG GAGAGGAAAA TATCTGATGC CAAGTCCCGT 480
 GAACTGGTGA TGTACATCTG GGCCCATGCA GTGGTGGCCA GTGTCCCTGT GTTTGCAGTA 540
 ACCAATGTGG CTGACATCTA TGCCACGTCC ACCTGCACGG AAGTCTGGAG CAACTCCTTG 600
 GGCCACCTGG TGTACGTTCT GGTGTATAAC ATCACCACGG TCATTGTGCC TGTGGTGGTG 660
 GTGTTCTCT TCTTGATACT GATCCGACGG GCCCTGAGTG CCAGCCAGAA GAAGAAGGTC 720
 30 ATCATAGCAG CGCTCCGGAC CCCACAGAAC ACCATCTCTA TTCCCTATGC CTCCCAGCGG 780
 GAGGCCGAGC TGAAAGCCAC CCTGCTCTCC ATGGTGATGG TCTTCATCTT GTGTAGCGTG 840
 CCCTATGCCA CCCTGGTCGT CTACCAGACT GTGCTCAATG TCCCTGACAC TTCCGTCTTC 900

TTGCTGCTCA CTGCTGTTTG GCTGCCCAAA GTCTCCCTGC TGGCAAACCC TGTTCCTTTT 960
 CTTACTGTGA ACAAATCTGT CCGCAAGTGC TTGATAGGGA CCCTGGTGCA ACTACACCAC 1020
 CGGTACAGTC GCCGTAATGT GGTCAGTACA GGGAGTGGCA TGGCTGAGGC CAGCCTGGAA 1080
 CCCAGCATAC GCTCGGGTAG CCAGCTCCTG GAGATGTTCC ACATTGGGCA GCAGCAGATC 1140
 5 TTTAAGCCCA CAGAGGATGA GGAAGAGAGT GAGGCCAAGT ACATTGGCTC AGCTGACTTC 1200
 CAGGCCAAGG AGATATTTAG CACCTGCCTG GAGGGAGAGC AGGGGCCACA GTTTGCGCCC 1260
 TCTGCCCCAC CCCTGAGCAC AGTGGACTCT GTATCCCAGG TGGCACCGGC AGCCCCCTGTG 1320
 GAACCTGAAA CATTCCCTGA TAAGTATTCC CTGCAGTTTG GCTTTGGGCC TTTTGAGTTG 1380
 CCTCCTCAGT GGCTCTCAGA GACCCGAAAC AGCAAGAAGC GGCTGCTTCC CCCCTTGGGC 1440
 10 AACACCCCAG AAGAGCTGAT CCAGACAAAG GTGCCCAAGG TAGGCAGGCT GGAGCGGAAG 1500
 ATGAGCAGAA ACAATAAAGT GAGCATTTTT CCAAAGGTGG ATTCCTAG 1548

(221) INFORMATION FOR SEQ ID NO:220:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 515 amino acids
 15 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

 (ii) MOLECULE TYPE: protein

 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:220:
 20 Met Gly His Asn Gly Ser Trp Ile Ser Pro Asn Ala Ser Glu Pro His
 1 5 10 15
 Asn Ala Ser Gly Ala Glu Ala Ala Gly Val Asn Arg Ser Ala Leu Gly
 20 25 30
 25 Glu Phe Gly Glu Ala Gln Leu Tyr Arg Gln Phe Thr Thr Thr Val Gln
 35 40 45
 Val Val Ile Phe Ile Gly Ser Leu Leu Gly Asn Phe Met Val Leu Trp
 50 55 60
 Ser Thr Cys Arg Thr Thr Val Phe Lys Ser Val Thr Asn Arg Phe Ile
 65 70 75 80
 30 Lys Asn Leu Ala Cys Ser Gly Ile Cys Ala Ser Leu Val Cys Val Pro
 85 90 95
 Phe Asp Ile Ile Leu Ser Thr Ser Pro His Cys Cys Trp Trp Ile Tyr
 100 105 110

	Thr Met Leu Phe Cys Lys Val Val Lys Phe Leu His Lys Val Phe Cys	
	115	120 125
	Ser Val Thr Ile Leu Ser Phe Pro Ala Ile Ala Leu Asp Arg Tyr Tyr	
	130	135 140
5	Ser Val Leu Tyr Pro Leu Glu Arg Lys Ile Ser Asp Ala Lys Ser Arg	
	145	150 155 160
	Glu Leu Val Met Tyr Ile Trp Ala His Ala Val Val Ala Ser Val Pro	
		165 170 175
10	Val Phe Ala Val Thr Asn Val Ala Asp Ile Tyr Ala Thr Ser Thr Cys	
		180 185 190
	Thr Glu Val Trp Ser Asn Ser Leu Gly His Leu Val Tyr Val Leu Val	
		195 200 205
	Tyr Asn Ile Thr Thr Val Ile Val Pro Val Val Val Val Phe Leu Phe	
		210 215 220
15	Leu Ile Leu Ile Arg Arg Ala Leu Ser Ala Ser Gln Lys Lys Lys Val	
		225 230 235 240
	Ile Ile Ala Ala Leu Arg Thr Pro Gln Asn Thr Ile Ser Ile Pro Tyr	
		245 250 255
20	Ala Ser Gln Arg Glu Ala Glu Leu Lys Ala Thr Leu Leu Ser Met Val	
		260 265 270
	Met Val Phe Ile Leu Cys Ser Val Pro Tyr Ala Thr Leu Val Val Tyr	
		275 280 285
	Gln Thr Val Leu Asn Val Pro Asp Thr Ser Val Phe Leu Leu Leu Thr	
		290 295 300
25	Ala Val Trp Leu Pro Lys Val Ser Leu Leu Ala Asn Pro Val Leu Phe	
		305 310 315 320
	Leu Thr Val Asn Lys Ser Val Arg Lys Cys Leu Ile Gly Thr Leu Val	
		325 330 335
30	Gln Leu His His Arg Tyr Ser Arg Arg Asn Val Val Ser Thr Gly Ser	
		340 345 350
	Gly Met Ala Glu Ala Ser Leu Glu Pro Ser Ile Arg Ser Gly Ser Gln	
		355 360 365
	Leu Leu Glu Met Phe His Ile Gly Gln Gln Gln Ile Phe Lys Pro Thr	
		370 375 380
35	Glu Asp Glu Glu Glu Ser Glu Ala Lys Tyr Ile Gly Ser Ala Asp Phe	
		385 390 395 400
	Gln Ala Lys Glu Ile Phe Ser Thr Cys Leu Glu Gly Glu Gln Gly Pro	

185

405 410 415
 Gln Phe Ala Pro Ser Ala Pro Pro Leu Ser Thr Val Asp Ser Val Ser
 420 425 430
 Gln Val Ala Pro Ala Ala Pro Val Glu Pro Glu Thr Phe Pro Asp Lys
 435 440 445
 Tyr Ser Leu Gln Phe Gly Phe Gly Pro Phe Glu Leu Pro Pro Gln Trp
 450 455 460
 Leu Ser Glu Thr Arg Asn Ser Lys Lys Arg Leu Leu Pro Pro Leu Gly
 465 470 475 480
 Asn Thr Pro Glu Glu Leu Ile Gln Thr Lys Val Pro Lys Val Gly Arg
 485 490 495
 Val Glu Arg Lys Met Ser Arg Asn Asn Lys Val Ser Ile Phe Pro Lys
 500 505 510
 Val Asp Ser
 515

(222) INFORMATION FOR SEQ ID NO:221:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1164 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:221:

ATGAATCGGC ACCATCTGCA GGATCACTTT CTGGAAATAG ACAAGAAGAA CTGCTGTGTG 60
 TTCCGAGATG ACTTCATTGC CAAGGTGTTG CCGCCGGTGT TGGGGCTGGA GTTTATCTTT 120
 GGGCTTCTGG GCAATGGCCT TGCCCTGTGG ATTTTCTGTT TCCACCTCAA GTCCTGGAAA 180
 TCCAGCCGGA TTTTCCTGTT CAACCTGGCA GTAGCTGACT TTCTACTGAT CATCTGCCTG 240
 CCGTTCGTGA TGGACTACTA TGTGCGGCGT TCAGACTGGA AGTTTGGGGA CATCCCTTGC 300
 CGGCTGGTGC TCTTCATGTT TGCCATGAAC CGCCAGGGCA GCATCATCTT CCTCACGGTG 360
 GTGGCGGTAG ACAGGTATTT CCGGGTGGTC CATCCCCACC ACGCCCTGAA CAAGATCTCC 420
 AATTGGACAG CAGCCATCAT CTCTTGCCTT CTGTGGGGCA TCACTGTTGG CCTAACAGTC 480
 CACCTCCTGA AGAAGAAGTT GCTGATCCAG AATGGCCCTG CAAATGTGTG CATCAGCTTC 540
 AGCATCTGCC ATACCTTCCG GTGGCAGGAA GCTATGTTCC TCCTGGAGTT CCTCCTGCCC 600

CTGGGCATCA TCCTGTTCTG CTCAGCCAGA ATTATCTGGA GCCTGCGGCA GAGACAAATG 660
 GACCGGCATG CCAAGATCAA GAGAGCCAAA ACCTTCATCA TGGTGGTGGC CATCGTCTTT 720
 GTCATCTGCT TCCTTCCCAG CGTGGTTGTG CGGATCCGCA TCTTCTGGCT CCTGCACACT 780
 TCGGGCACGC AGAATTGTGA AGTGTACCGC TCGGTGGACC TGGCGTTCTT TATCACTCTC 840
 5 AGCTTCACCT ACATGAACAG CATGCTGGAC CCCGTGGTGT ACTACTTCTC CAGCCCATCC 900
 TTTCCCAACT TCTTCTCCAC TTTGATCAAC CGCTGCCTCC AGAGGAAGAT GACAGGTGAG 960
 CCAGATAATA ACCGCAGCAC GAGCGTCGAG CTCACAGGGG ACCCCAACAA AACCAGAGGC 1020
 GCTCCAGAGG CGTTAATGGC CAACTCCGGT GAGCCATGGA GCCCCTCTTA TCTGGGCCCA 1080
 ACCTCAAATA ACCATTCCAA GAAGGGACAT TGTACCAAG AACCAGCATC TCTGGAGAAA 1140
 10 CAGTTGGGCT GTTGCATCGA GTAA 1164

(223) INFORMATION FOR SEQ ID NO:222:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 387 amino acids
 (B) TYPE: amino acid
 15 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:222:

20 Met Asn Arg His His Leu Gln Asp His Phe Leu Glu Ile Asp Lys Lys
 1 5 10 15
 Asn Cys Cys Val Phe Arg Asp Asp Phe Ile Ala Lys Val Leu Pro Pro
 20 25 30
 Val Leu Gly Leu Glu Phe Ile Phe Gly Leu Leu Gly Asn Gly Leu Ala
 35 40 45
 25 Leu Trp Ile Phe Cys Phe His Leu Lys Ser Trp Lys Ser Ser Arg Ile
 50 55 60
 Phe Leu Phe Asn Leu Ala Val Ala Asp Phe Leu Leu Ile Ile Cys Leu
 65 70 75 80
 30 Pro Phe Val Met Asp Tyr Tyr Val Arg Arg Ser Asp Trp Lys Phe Gly
 85 90 95
 Asp Ile Pro Cys Arg Leu Val Leu Phe Met Phe Ala Met Asn Arg Gln
 100 105 110
 Gly Ser Ile Ile Phe Leu Thr Val Val Ala Val Asp Arg Tyr Phe Arg
 115 120 125

187

	Val	Val	His	Pro	His	His	Ala	Leu	Asn	Lys	Ile	Ser	Asn	Trp	Thr	Ala	
	130						135					140					
	Ala	Ile	Ile	Ser	Cys	Leu	Leu	Trp	Gly	Ile	Thr	Val	Gly	Leu	Thr	Val	
	145					150					155					160	
5	His	Leu	Leu	Lys	Lys	Lys	Leu	Leu	Ile	Gln	Asn	Gly	Pro	Ala	Asn	Val	
					165					170					175		
	Cys	Ile	Ser	Phe	Ser	Ile	Cys	His	Thr	Phe	Arg	Trp	His	Glu	Ala	Met	
				180					185					190			
10	Phe	Leu	Leu	Glu	Phe	Leu	Leu	Pro	Leu	Gly	Ile	Ile	Leu	Phe	Cys	Ser	
		195						200					205				
	Ala	Arg	Ile	Ile	Trp	Ser	Leu	Arg	Gln	Arg	Gln	Met	Asp	Arg	His	Ala	
	210						215					220					
	Lys	Ile	Lys	Arg	Ala	Lys	Thr	Phe	Ile	Met	Val	Val	Ala	Ile	Val	Phe	
	225					230					235					240	
15	Val	Ile	Cys	Phe	Leu	Pro	Ser	Val	Val	Val	Arg	Ile	Arg	Ile	Phe	Trp	
					245					250					255		
	Leu	Leu	His	Thr	Ser	Gly	Thr	Gln	Asn	Cys	Glu	Val	Tyr	Arg	Ser	Val	
				260					265					270			
20	Asp	Leu	Ala	Phe	Phe	Ile	Thr	Leu	Ser	Phe	Thr	Tyr	Met	Asn	Ser	Met	
		275						280					285				
	Leu	Asp	Pro	Val	Val	Tyr	Tyr	Phe	Ser	Ser	Pro	Ser	Phe	Pro	Asn	Phe	
		290					295					300					
	Phe	Ser	Thr	Leu	Ile	Asn	Arg	Cys	Leu	Gln	Arg	Lys	Met	Thr	Gly	Glu	
	305					310					315					320	
25	Pro	Asp	Asn	Asn	Arg	Ser	Thr	Ser	Val	Glu	Leu	Thr	Gly	Asp	Pro	Asn	
					325					330					335		
	Lys	Thr	Arg	Gly	Ala	Pro	Glu	Ala	Leu	Met	Ala	Asn	Ser	Gly	Glu	Pro	
				340					345					350			
30	Trp	Ser	Pro	Ser	Tyr	Leu	Gly	Pro	Thr	Ser	Asn	Asn	His	Ser	Lys	Lys	
		355						360					365				
	Gly	His	Cys	His	Gln	Glu	Pro	Ala	Ser	Leu	Glu	Lys	Gln	Leu	Gly	Cys	
		370					375					380					
	Cys	Ile	Glu														
	385																

35 (224) INFORMATION FOR SEQ ID NO:223:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1212 base pairs

(B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: DNA (genomic)

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:223:

	ATGGCTTGCA ATGGCAGTGC GGCCAGGGGG CACTTTGACC CTGAGGACTT GAACCTGACT	60
	GACGAGGCAC TGAGACTCAA GTACCTGGGG CCCCAGCAGA CAGAGCTGTT CATGCCCATC	120
	TGTGCCACAT ACCTGCTGAT CTTCTGTTGG GGCCTGTGG GCAATGGGCT GACCTGTCTG	180
	GTCATCCTGC GCCACAAGGC CATGCGCACG CCTACCAACT ACTACCTCTT CAGCCTGGCC	240
10	GTGTCCGACC TGCTGGTGCT GCTGGTGGGC CTGCCCCCTGG AGCTCTATGA GATGTGGCAC	300
	AACTACCCCT TCCTGCTGGG CGTTGGTGGC TGCTATTTCC GCACGCTACT GTTTGAGATG	360
	GTCTGCCTGG CCTCAGTGCT CAACGTCCT GGCCTGAGCG TGGAACGCTA TGTGGCCGTG	420
	GTGCACCCAC TCCAGGCCAG GTCCATGGTG ACGCGGGCCC ATGTGCGCCG AGTGCTTGGG	480
	GCCGTCTGGG GTCTTGCCAT GCTCTGCTCC CTGCCCCAACA CCAGCCTGCA CGGCATCCGG	540
15	CAGCTGCACG TGCCCTGCCG GGGCCAGTG CCAGACTCAG CTGTTTGCAT GCTGGTCCGC	600
	CCACGGGCCC TCTACAACAT GGTAGTGCG ACCACCGCGC TGCTCTTCTT CTGCCTGCCC	660
	ATGGCCATCA TGAGCGTGCT CTACCTGCTC ATTGGGCTGC GACTGCGGCG GGAGAGGCTG	720
	CTGCTCATGC AGGAGGCCAA GGGCAGGGGC TCTGCAGCAG CCAGGTCCAG ATACACCTGC	780
	AGGCTCCAGC AGCAGGATCG GGGCCGGAGA CAAGTGAAGA AGATGCTGTT TGTCTGGTC	840
20	GTGGTGTTTG GCATCTGCTG GGCCCCGTTT CACGCCGACC GCGTCATGTG GAGCGTCGTG	900
	TCACAGTGGA CAGATGGCCT GCACCTGGCC TTCCAGCAGC TGCACGTCAT CTCCGGCATC	960
	TTCTTCTACC TGGGCTCGGC GGCCAACCCC GTGCTCTATA GCCTCATGTC CAGCCGCTTC	1020
	CGAGAGACCT TCCAGGAGGC CCTGTGCCTC GGGGCCTGCT GCCATCGCCT CAGACCCCGC	1080
	CACAGCTCCC ACAGCCTCAG CAGGATGACC ACAGGCAGCA CCCTGTGTGA TGTGGGCTCC	1140
25	CTGGGCAGCT GGGTCCACCC CCTGGCTGGG AACGATGGCC CAGAGGCGCA GCAAGAGACC	1200
	GATCCATCCT GA	1212

(225) INFORMATION FOR SEQ ID NO:224:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 403 amino acids
 (B) TYPE: amino acid

30

189

(C) STRANDEDNESS:

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:224:

```

5   Met Ala Cys Asn Gly Ser Ala Ala Arg Gly His Phe Asp Pro Glu Asp
    1             5             10             15

    Leu Asn Leu Thr Asp Glu Ala Leu Arg Leu Lys Tyr Leu Gly Pro Gln
           20             25             30

10   Gln Thr Glu Leu Phe Met Pro Ile Cys Ala Thr Tyr Leu Leu Ile Phe
           35             40             45

    Val Val Gly Ala Val Gly Asn Gly Leu Thr Cys Leu Val Ile Leu Arg
        50             55             60

    His Lys Ala Met Arg Thr Pro Thr Asn Tyr Tyr Leu Phe Ser Leu Ala
    65             70             75             80

15   Val Ser Asp Leu Leu Val Leu Leu Val Gly Leu Pro Leu Glu Leu Tyr
           85             90             95

    Glu Met Trp His Asn Tyr Pro Phe Leu Leu Gly Val Gly Gly Cys Tyr
           100            105            110

20   Phe Arg Thr Leu Leu Phe Glu Met Val Cys Leu Ala Ser Val Leu Asn
           115            120            125

    Val Thr Ala Leu Ser Val Glu Arg Tyr Val Ala Val Val His Pro Leu
           130            135            140

    Gln Ala Arg Ser Met Val Thr Arg Ala His Val Arg Arg Val Leu Gly
    145            150            155            160

25   Ala Val Trp Gly Leu Ala Met Leu Cys Ser Leu Pro Asn Thr Ser Leu
           165            170            175

    His Gly Ile Arg Gln Leu His Val Pro Cys Arg Gly Pro Val Pro Asp
           180            185            190

    Ser Ala Val Cys Met Leu Val Arg Pro Arg Ala Leu Tyr Asn Met Val
    195            200            205

30   Val Gln Thr Thr Ala Leu Leu Phe Phe Cys Leu Pro Met Ala Ile Met
           210            215            220

    Ser Val Leu Tyr Leu Leu Ile Gly Leu Arg Leu Arg Arg Glu Arg Leu
    225            230            235            240

35   Leu Leu Met Gln Glu Ala Lys Gly Arg Gly Ser Ala Ala Ala Arg Ser
           245            250            255

```

Arg Tyr Thr Cys Arg Leu Gln Gln His Asp Arg Gly Arg Arg Gln Val
 260 265 270
 Lys Lys Met Leu Phe Val Leu Val Val Val Phe Gly Ile Cys Trp Ala
 275 280 285
 5 Pro Phe His Ala Asp Arg Val Met Trp Ser Val Val Ser Gln Trp Thr
 290 295 300
 Asp Gly Leu His Leu Ala Phe Gln His Val His Val Ile Ser Gly Ile
 305 310 315 320
 10 Phe Phe Tyr Leu Gly Ser Ala Ala Asn Pro Val Leu Tyr Ser Leu Met
 325 330 335
 Ser Ser Arg Phe Arg Glu Thr Phe Gln Glu Ala Leu Cys Leu Gly Ala
 340 345 350
 Cys Cys His Arg Leu Arg Pro Arg His Ser Ser His Ser Leu Ser Arg
 355 360 365
 15 Met Thr Thr Gly Ser Thr Leu Cys Asp Val Gly Ser Leu Gly Ser Trp
 370 375 380
 Val His Pro Leu Ala Gly Asn Asp Gly Pro Glu Ala Gln Gln Glu Thr
 385 390 395 400
 20 Asp Pro Ser

(226) INFORMATION FOR SEQ ID NO:225:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1098 base pairs
 (B) TYPE: nucleic acid
 25 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:225:

ATGGGGAACA TCACTGCAGA CAACTCCTCG ATGAGCTGTA CCATCGACCA TACCATCCAC 60
 30 CAGACGCTGG CCCCGTGGT CTATGTTACC GTGCTGGTGG TGGGCTTCCC GGCCAACTGC 120
 CTGTCCCTCT ACTTCGGCTA CCTGCAGATC AAGGCCCGGA ACGAGCTGGG CGTGTACCTG 180
 TGCAACCTGA CGGTGGCCGA CCTCTTCTAC ATCTGCTCGC TGCCCTTCTG GCTGCAGTAC 240
 GTGCTGCAGC ACGACAACTG GTCTCACGGC GACCTGTCCT GCCAGGTGTG CGGCATCCTC 300
 CTGTACGAGA ACATCTACAT CAGCGTGGGC TTCCTCTGCT GCATCTCCGT GGACCGCTAC 360
 35 CTGGCTGTGG CCCATCCCTT CCGCTTCCAC CAGTTCGGGA CCCTGAAGGC GGCCGTCGGC 420

191

GTCAGCGTGG TCATCTGGGC CAAGGAGCTG CTGACCAGCA TCTACTTCCT GATGCACGAG 480
 GAGGTCATCG AGGACGAGAA CCAGCACCGC GTGTGCTTTG AGCACTACCC CATCCAGGCA 540
 TGGCAGCGCG CCATCAACTA CTACCGCTTC CTGGTGGGCT TCCTCTTCCC CATCTGCCTG 600
 CTGCTGGCGT CCTACCAGGG CATCCTGCGC GCCGTGCGCC GGAGCCACGG CACCCAGAAG 660
 5 AGCCGCAAGG ACCAGATCAA GCGGCTGGTG CTCAGCACCG TGGTCATCTT CCTGGCCTGC 720
 TTCCTGCCCT ACCACGTGTT GCTGCTGGTG CGCAGCGTCT GGGAGGCCAG CTGCGACTTC 780
 GCCAAGGGCG TTTTCAACGC CTACCACTTC TCCCTCCTGC TCACCAGCTT CAACTGCGTC 840
 GCCGACCCCG TGCTCTACTG CTTCGTCAGC GAGACCACCC ACCGGGACCT GGCCCGCCTC 900
 CGCGGGGCCT GCCTGGCCTT CCTCACCTGC TCCAGGACCG GCCGGGCCAG GGAGGCCTAC 960
 10 CCGCTGGSTG CCCCCGAGGC CTCCGGGAAA AGCGGGGCCC AGGGTGAGGA GCCCAGAGTG 1020
 TTGACCAAGC TCCACCCGGC CTTCCAGACC CCTAACTCGC CAGGGTCGGG CGGGTTCCCC 1080
 ACGGGCAGGT TGGCCTAG 1098

(227) INFORMATION FOR SEQ ID NO:226:

15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 365 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:226:

Met Gly Asn Ile Thr Ala Asp Asn Ser Ser Met Ser Cys Thr Ile Asp
 1 5 10 15

His Thr Ile His Gln Thr Leu Ala Pro Val Val Tyr Val Thr Val Leu
 20 25 30

25 Val Val Gly Phe Pro Ala Asn Cys Leu Ser Leu Tyr Phe Gly Tyr Leu
 35 40 45

Gln Ile Lys Ala Arg Asn Glu Leu Gly Val Tyr Leu Cys Asn Leu Thr
 50 55 60

30 Val Ala Asp Leu Phe Tyr Ile Cys Ser Leu Pro Phe Trp Leu Gln Tyr
 65 70 75 80

Val Leu Gln His Asp Asn Trp Ser His Gly Asp Leu Ser Cys Gln Val
 85 90 95

Cys Gly Ile Leu Leu Tyr Glu Asn Ile Tyr Ile Ser Val Gly Phe Leu

(228) INFORMATION FOR SEQ ID NO:227:

BRISBANE (WA) 0022120A: JA>

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:227:

5	ATGGATATTC TTTGTGAAGA AAATACTTCT TTGAGCTCAA CTACGAACTC CCTAATGCAA	60
	TTAAATGATG ACAACAGGCT CTACAGTAAT GACTTTAACT CCGGAGAAGC TAACACTTCT	120
	GATGCATTTA ACTGGACAGT CGACTCTGAA AATCGAACCA ACCTTTCCTG TGAAGGGTGC	180
	CTCTCACCGT CGTGTCTCTC CTTACTTCAT CTCCAGGAAA AAAACTGGTC TGCTTTACTG	240
	ACAGCCGTAG TGATTATTCT AACTATTGCT GGAAACATAC TCGTCATCAT GGCAGTGTCC	300
10	CTAGAGAAAA AGCTGCAGAA TGCCACCAAC TATTTCTCTGA TGTCACTTGC CATAGCTGAT	360
	ATGCTGCTGG GTTTCCTTGT CATGCCCGTG TCCATGTAA CCATCCTGTA TGGGTACCGG	420
	TGGCCTCTGC CGAGCAAGCT TTGTGCAGTC TGGATTTACC TGGACGTGCT CTTCTCCACG	480
	GCCTCCATCA TGCACCTCTG CGCCATCTCG CTGGACCGCT ACGTCGCCAT CCAGAATCCC	540
	ATCCACCACA GCCGCTTCAA CTCCAGAACT AAGGCATTTT TGAAAATCAT TGCTGTTTGG	600
15	ACCATATCAG TAGGTATATC CATGCCAATA CCAGTCTTTG GGCTACAGGA CGATTCTGAAG	660
	GTCTTTAAGG AGGGGAGTTG CTTACTCGCC GATGATAACT TTGTCCTGAT CGGCTCTTTT	720
	GTGTCATTTT TCATTCCCTT AACCATCATG GTGATCACCT ACTTTCTAAC TATCAAGTCA	780
	CTCCAGAAAG AAGCTACTTT GTGTGTAAGT GATCTTGGCA CACGGGCCAA ATTAGCTTCT	840
	TTCAGCTTCC TCCCTCAGAG TTCTTTGTCT TCAGAAAAGC TCTTCCAGCG GTCGATCCAT	900
20	AGGGAGCCAG GGTCTACAC AGGCAGGAGG ACTATGCAGT CCATCAGCAA TGAGCAAAAAG	960
	GCAAAGAAGG TGCTGGGCAT CGTCTTCTTC CTGTTTGTGG TGATGTGGTG CCCTTCTTC	1020
	ATCACAAACA TCATGGCCGT CATCTGCAAA GAGTCCTGCA ATGAGGATGT CATTGGGGCC	1080
	CTGCTCAATG TGTTTGTGTTG GATCGGTTAT CTCTCTTCAG CAGTCAACCC ACTAGTCTAC	1140
	ACACTGTTCA ACAAGACCTA TAGGTCAGCC TTTTCACGGT ATATTCAGTG TCAGTACAAG	1200
25	GAAAACAAAA AACCATTGCA GTTAATTTTA GTGAACACAA TACCGGCTTT GGCCTACAAG	1260
	TCTAGCCAAC TTCAAATGGG ACAAAAAAAG AATTCAAAGC AAGATGCCAA GACAACAGAT	1320
	AATGACTGCT CAATGGTTGC TCTAGGAAAG CAGTATTCTG AAGAGGCTTC TAAAGACAAT	1380
	AGCGACGGAG TGAATGAAAA GGTGAGCTGT GTGTGA	1416

(229) INFORMATION FOR SEQ ID NO:228:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 470 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:228:

```

10  1  Met Asp Ile Leu Cys Glu Glu Asn Thr Ser Leu Ser Ser Thr Thr Asn
      5              10              15
    Ser Leu Met Gln Leu Asn Asp Asp Asn Arg Leu Tyr Ser Asn Asp Phe
      20              25              30

    Asn Ser Gly Glu Ala Asn Thr Ser Asp Ala Phe Asn Trp Thr Val Asp
      35              40              45

15  Ser Glu Asn Arg Thr Asn Leu Ser Cys Glu Gly Cys Leu Ser Pro Ser
      50              55              60

    Cys Leu Ser Leu Leu His Leu Gln Glu Lys Asn Trp Ser Ala Leu Leu
      65              70              75              80

20  Thr Ala Val Val Ile Ile Leu Thr Ile Ala Gly Asn Ile Leu Val Ile
      85              90              95

    Met Ala Val Ser Leu Glu Lys Lys Leu Gln Asn Ala Thr Asn Tyr Phe
      100             105             110

    Leu Met Ser Leu Ala Ile Ala Asp Met Leu Leu Gly Phe Leu Val Met
      115             120             125

25  Pro Val Ser Met Leu Thr Ile Leu Tyr Gly Tyr Arg Trp Pro Leu Pro
      130             135             140

    Ser Lys Leu Cys Ala Val Trp Ile Tyr Leu Asp Val Leu Phe Ser Thr
      145             150             155             160

30  Ala Ser Ile Met His Leu Cys Ala Ile Ser Leu Asp Arg Tyr Val Ala
      165             170             175

    Ile Gln Asn Pro Ile His His Ser Arg Phe Asn Ser Arg Thr Lys Ala
      180             185             190

    Phe Leu Lys Ile Ile Ala Val Trp Thr Ile Ser Val Gly Ile Ser Met
      195             200             205

35  Pro Ile Pro Val Phe Gly Leu Gln Asp Asp Ser Lys Val Phe Lys Glu
      210             215             220

    Gly Ser Cys Leu Leu Ala Asp Asp Asn Phe Val Leu Ile Gly Ser Phe
      225             230             235             240

```

195

	Val	Ser	Phe	Phe	Ile	Pro	Leu	Thr	Ile	Met	Val	Ile	Thr	Tyr	Phe	Leu	
					245					250					255		
	Thr	Ile	Lys	Ser	Leu	Gln	Lys	Glu	Ala	Thr	Leu	Cys	Val	Ser	Asp	Leu	
				260					265					270			
5	Gly	Thr	Arg	Ala	Lys	Leu	Ala	Ser	Phe	Ser	Phe	Leu	Pro	Gln	Ser	Ser	
			275					280					285				
	Leu	Ser	Ser	Glu	Lys	Leu	Phe	Gln	Arg	Ser	Ile	His	Arg	Glu	Pro	Gly	
		290					295					300					
10	Ser	Tyr	Thr	Gly	Arg	Arg	Thr	Met	Gln	Ser	Ile	Ser	Asn	Glu	Gln	Lys	
	305					310					315					320	
	Ala	Lys	Lys	Val	Leu	Gly	Ile	Val	Phe	Phe	Leu	Phe	Val	Val	Met	Trp	
					325					330					335		
	Cys	Pro	Phe	Phe	Ile	Thr	Asn	Ile	Met	Ala	Val	Ile	Cys	Lys	Glu	Ser	
				340					345					350			
15	Cys	Asn	Glu	Asp	Val	Ile	Gly	Ala	Leu	Leu	Asn	Val	Phe	Val	Trp	Ile	
			355					360					365				
	Gly	Tyr	Leu	Ser	Ser	Ala	Val	Asn	Pro	Leu	Val	Tyr	Thr	Leu	Phe	Asn	
		370					375					380					
20	Lys	Thr	Tyr	Arg	Ser	Ala	Phe	Ser	Arg	Tyr	Ile	Gln	Cys	Gln	Tyr	Lys	
	385					390					395					400	
	Glu	Asn	Lys	Lys	Pro	Leu	Gln	Leu	Ile	Leu	Val	Asn	Thr	Ile	Pro	Ala	
					405					410					415		
	Leu	Ala	Tyr	Lys	Ser	Ser	Gln	Leu	Gln	Met	Gly	Gln	Lys	Lys	Asn	Ser	
				420					425					430			
25	Lys	Gln	Asp	Ala	Lys	Thr	Thr	Asp	Asn	Asp	Cys	Ser	Met	Val	Ala	Leu	
			435					440					445				
	Gly	Lys	Gln	Tyr	Ser	Glu	Glu	Ala	Ser	Lys	Asp	Asn	Ser	Asp	Gly	Val	
		450					455					460					
30	Asn	Glu	Lys	Val	Ser	Cys	Val										
	465					470											

(230) INFORMATION FOR SEQ ID NO:229:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1377 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:229:

ATGGTGAACC TGAGGAATGC GGTGCATTCA TTCCTTGTGC ACCTAATTGG CCTATTGGTT 60
 TGGCAATGTG ATATTTCTGT GAGCCCAGTA GCAGCTATAG TAACTGACAT TTTCAATACC 120
 TCCGATGGTG GACGCTTCAA ATTCCCAGAC GGGGTACAAA ACTGGCCAGC ACTTTCAATC 180
 5 GTCATCATAA TAATCATGAC AATAGGTGGC AACATCCTTG TGATCATGGC AGTAAGCATG 240
 GAAAAGAAAC TGCACAATGC CACCAATTAC TTCTTAATGT CCCTAGCCAT TGCTGATATG 300
 CTAGTGGGAC TACTTGTCAT GCCCTGTCT CTCCTGGCAA TCCTTTATGA TTATGTCTGG 360
 CCACTACCTA GATATTTGTG CCCCGTCTGG ATTTCTTTAG ATGTTTTATT TTCAACAGCG 420
 TCCATCATGC ACCTCTGCGC TATATCGCTG GATCGGTATG TAGCAATACG TAATCCTATT 480
 10 GAGCATAGCC GTTTC AATTC GCGGACTAAG GCCATCATGA AGATTGCTAT TGTTTGGGCA 540
 ATTTCTATAG GTGTATCAGT TCCTATCCCT GTGATTGGAC TGAGGGACGA AGAAAAGGTG 600
 TTCGTGAACA ACACGACGTG CGTGCTCAAC GACCCAAATT TCGTTCTTAT TGGGTCCTTC 660
 GTAGCTTTCT TCATACCGCT GACGATTATG GTGATTACGT ATTGCCTGAC CATCTACGTT 720
 CTGCGCCGAC AAGCTTTGAT GTTACTGCAC GGCCACACCG AGGAACCGCC TGGACTAAGT 780
 15 CTGGATTTCC TGAAGTGCTG CAAGAGGAAT ACGGCCGAGG AAGAGAACTC TGCAAACCCT 840
 AACCAAGACC AGAACGCACG CCGAAGAAAG AAGAAGGAGA GACGTCCTAG GGGCACCATG 900
 CAGGCTATCA ACAATGAAAG AAAAGCTAAG AAAGTCCTTG GGATTGTTTT CTTTGTGTTT 960
 CTGATCATGT GGTGCCCATT TTTCATTACC AATATTCTGT CTGTTCTTTG TGAGAAGTCC 1020
 TGTAACCAAA AGCTCATGGA AAAGCTTCTG AATGTGTTTG TTTGGATTGG CTATGTTTGT 1080
 20 TCAGGAATCA ATCCTCTGGT GTATACTCTG TTCAACAAAA TTTACCGAAG GGCATTCTCC 1140
 AACTATTTGC GTTGCAATTA TAAGGTAGAG AAAAAGCCTC CTGTCAGGCA GATTCCAAGA 1200
 GTTGCCGCCA CTGCTTTGTC TGGGAGGGAG CTTAATGTTA ACATTTATCG GCATACCAAT 1260
 GAACCGGTGA TCGAGAAAGC CAGTGACAAT GAGCCCGGTA TAGAGATGCA AGTTGAGAAT 1320
 TTAGAGTTAC CAGTAAATCC CTCCAGTGTG GTTAGCGAAA GGATTAGCAG TGTGTGA 1377

25 (231) INFORMATION FOR SEQ ID NO:230:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 458 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:

197

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:230:

5 Met Val Asn Leu Arg Asn Ala Val His Ser Phe Leu Val His Leu Ile
 1 5 10 15
 Gly Leu Leu Val Trp Gln Cys Asp Ile Ser Val Ser Pro Val Ala Ala
 20 25 30
 Ile Val Thr Asp Ile Phe Asn Thr Ser Asp Gly Gly Arg Phe Lys Phe
 35 40 45
 10 Pro Asp Gly Val Gln Asn Trp Pro Ala Leu Ser Ile Val Ile Ile Ile
 50 55 60
 Ile Met Thr Ile Gly Gly Asn Ile Leu Val Ile Met Ala Val Ser Met
 65 70 75 80
 15 Glu Lys Lys Leu His Asn Ala Thr Asn Tyr Phe Leu Met Ser Leu Ala
 85 90 95
 Ile Ala Asp Met Leu Val Gly Leu Leu Val Met Pro Leu Ser Leu Leu
 100 105 110
 Ala Ile Leu Tyr Asp Tyr Val Trp Pro Leu Pro Arg Tyr Leu Cys Pro
 115 120 125
 20 Val Trp Ile Ser Leu Asp Val Leu Phe Ser Thr Ala Ser Ile Met His
 130 135 140
 Leu Cys Ala Ile Ser Leu Asp Arg Tyr Val Ala Ile Arg Asn Pro Ile
 145 150 155 160
 25 Glu His Ser Arg Phe Asn Ser Arg Thr Lys Ala Ile Met Lys Ile Ala
 165 170 175
 Ile Val Trp Ala Ile Ser Ile Gly Val Ser Val Pro Ile Pro Val Ile
 180 185 190
 Gly Leu Arg Asp Glu Glu Lys Val Phe Val Asn Asn Thr Thr Cys Val
 195 200 205
 30 Leu Asn Asp Pro Asn Phe Val Leu Ile Gly Ser Phe Val Ala Phe Phe
 210 215 220
 Ile Pro Leu Thr Ile Met Val Ile Thr Tyr Cys Leu Thr Ile Tyr Val
 225 230 235 240
 35 Leu Arg Arg Gln Ala Leu Met Leu Leu His Gly His Thr Glu Glu Pro
 245 250 255
 Pro Gly Leu Ser Leu Asp Phe Leu Lys Cys Cys Lys Arg Asn Thr Ala

	260	265	270
	Glu Glu Glu Asn Ser Ala Asn Pro Asn Gln Asp Gln Asn Ala Arg Arg		
	275	280	285
5	Arg Lys Lys Lys Glu Arg Arg Pro Arg Gly Thr Met Gln Ala Ile Asn		
	290	295	300
	Asn Glu Arg Lys Ala Lys Lys Val Leu Gly Ile Val Phe Phe Val Phe		
	305	310	315 320
	Leu Ile Met Trp Cys Pro Phe Phe Ile Thr Asn Ile Leu Ser Val Leu		
	325	330	335
10	Cys Glu Lys Ser Cys Asn Gln Lys Leu Met Glu Lys Leu Leu Asn Val		
	340	345	350
	Phe Val Trp Ile Gly Tyr Val Cys Ser Gly Ile Asn Pro Leu Val Tyr		
	355	360	365
15	Thr Leu Phe Asn Lys Ile Tyr Arg Arg Ala Phe Ser Asn Tyr Leu Arg		
	370	375	380
	Cys Asn Tyr Lys Val Glu Lys Lys Pro Pro Val Arg Gln Ile Pro Arg		
	385	390	395 400
	Val Ala Ala Thr Ala Leu Ser Gly Arg Glu Leu Asn Val Asn Ile Tyr		
	405	410	415
20	Arg His Thr Asn Glu Pro Val Ile Glu Lys Ala Ser Asp Asn Glu Pro		
	420	425	430
	Gly Ile Glu Met Gln Val Glu Asn Leu Glu Leu Pro Val Asn Pro Ser		
	435	440	445
25	Ser Val Val Ser Glu Arg Ile Ser Ser Val		
	450	455	

(232) INFORMATION FOR SEQ ID NO:231:

30	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 1068 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:231:	
	ATGGATCAGT TCCCTGAATC AGTGACAGAA AACTTTGAGT ACGATGATTT GGCTGAGGCC	60
35	TGTTATATTG GGGACATCGT GGTCTTTGGG ACTGTGTTCC TGTCCATATT CTACTCCGTC	120
	ATCTTTGCCA TTGGCCTGGT GGGAAATTTG TTGGTAGTGT TTGCCCTCAC CAACAGCAAG	180

199

AAGCCCAAGA GTGTCACCGA CATTACCTC CTGAACCTGG CCTTGTCTGA TCTGCTGTTT 240
 GTAGCCACTT TGCCCTTCTG GACTCACTAT TTGATAAATG AAAAGGGCCT CCACAATGCC 300
 ATGTGCAAAT TCACTACCGC CTTCTTCTTC ATCGGCTTTT TTGAAGCAT ATTCTTCATC 360
 ACCGTCATCA GCATTGATAG GTACCTGGCC ATCGTCCTGG CCGCCAACTC CATGAACAAC 420
 5 CGGACCGTGC AGCATGGCGT CACCATCAGC CTAGGCGTCT GGGCAGCAGC CATTTTGGTG 480
 GCAGCACCCC AGTTCATGTT CACAAAGCAG AAAGAAAATG AATGCCTTGG TGACTIONCCC 540
 GAGGTCCTCC AGGAAATCTG GCCCGTGCTC CGCAATGTGG AAACAAATTT TCTTGGCTTC 600
 CTACTCCCCC TGCTCATTAT GAGTTATTGC TACTTCAGAA TCATCCAGAC GCTGTTTTCC 660
 TGCAAGAACC ACAAGAAAGC CAAAGCCAAG AAAGTATGCC TTCTGGTGGT CATCGTGTTC 720
 10 TTCCTCTTCT GGACACCCTA CAACGTTATG ATTTTCCTGG AGACGCTTAA GCTCTATGAC 780
 TTCTTTCCCA GTTGTGACAT GAGGAAGGAT CTGAGGCTGG CCCTCAGTGT GACTGAGACG 840
 GTTGCATTTA GCCATTGTTG CCTGAATCCT CTCATCTATG CATTGCTGG GGAGAAGTTC 900
 AGAAGATACC TTTACCACCT GTATGGGAAA TGCCTGGCTG TCCTGTGTGG GCGCTCAGTC 960
 CACGTTGATT TCTCCTCATC TGAATCACAA AGGAGCAGGC ATGGAAGTGT TCTGAGCAGC 1020
 15 AATTTTACTT ACCACACGAG TGATGGAGAT GCATTGCTCC TTCTCTGA 1068

(233) INFORMATION FOR SEQ ID NO:232:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 355 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:232:

25 Met Asp Gln Phe Pro Glu Ser Val Thr Glu Asn Phe Glu Tyr Asp Asp
 1 5 10 15
 Leu Ala Glu Ala Cys Tyr Ile Gly Asp Ile Val Val Phe Gly Thr Val
 20 25 30
 Phe Leu Ser Ile Phe Tyr Ser Val Ile Phe Ala Ile Gly Leu Val Gly
 35 40 45
 30 Asn Leu Leu Val Val Phe Ala Leu Thr Asn Ser Lys Lys Pro Lys Ser
 50 55 60
 Val Thr Asp Ile Tyr Leu Leu Asn Leu Ala Leu Ser Asp Leu Leu Phe

	65		70		75		80									
	Val	Ala	Thr	Leu	Pro	Phe	Trp	Thr	His	Tyr	Leu	Ile	Asn	Glu	Lys	Gly
				85						90					95	
5	Leu	His	Asn	Ala	Met	Cys	Lys	Phe	Thr	Thr	Ala	Phe	Phe	Phe	Ile	Gly
				100					105						110	
	Phe	Phe	Gly	Ser	Ile	Phe	Phe	Ile	Thr	Val	Ile	Ser	Ile	Asp	Arg	Tyr
			115						120					125		
	Leu	Ala	Ile	Val	Leu	Ala	Ala	Asn	Ser	Met	Asn	Asn	Arg	Thr	Val	Gln
		130						135					140			
10	His	Gly	Val	Thr	Ile	Ser	Leu	Gly	Val	Trp	Ala	Ala	Ala	Ile	Leu	Val
	145					150					155				160	
	Ala	Ala	Pro	Gln	Phe	Met	Phe	Thr	Lys	Gln	Lys	Glu	Asn	Glu	Cys	Leu
					165					170					175	
15	Gly	Asp	Tyr	Pro	Glu	Val	Leu	Gln	Glu	Ile	Trp	Pro	Val	Leu	Arg	Asn
			180						185					190		
	Val	Glu	Thr	Asn	Phe	Leu	Gly	Phe	Leu	Leu	Pro	Leu	Leu	Ile	Met	Ser
		195						200						205		
	Tyr	Cys	Tyr	Phe	Arg	Ile	Ile	Gln	Thr	Leu	Phe	Ser	Cys	Lys	Asn	His
	210					215						220				
20	Lys	Lys	Ala	Lys	Ala	Lys	Lys	Leu	Ile	Leu	Leu	Val	Val	Ile	Val	Phe
	225					230					235				240	
	Phe	Leu	Phe	Trp	Thr	Pro	Tyr	Asn	Val	Met	Ile	Phe	Leu	Glu	Thr	Leu
				245						250					255	
25	Lys	Leu	Tyr	Asp	Phe	Phe	Pro	Ser	Cys	Asp	Met	Arg	Lys	Asp	Leu	Arg
			260						265					270		
	Leu	Ala	Leu	Ser	Val	Thr	Glu	Thr	Val	Ala	Phe	Ser	His	Cys	Cys	Leu
		275						280					285			
	Asn	Pro	Leu	Ile	Tyr	Ala	Phe	Ala	Gly	Glu	Lys	Phe	Arg	Arg	Tyr	Leu
	290					295						300				
30	Tyr	His	Leu	Tyr	Gly	Lys	Cys	Leu	Ala	Val	Leu	Cys	Gly	Arg	Ser	Val
	305					310					315				320	
	His	Val	Asp	Phe	Ser	Ser	Ser	Glu	Ser	Gln	Arg	Ser	Arg	His	Gly	Ser
				325						330				335		
35	Val	Leu	Ser	Ser	Asn	Phe	Thr	Tyr	His	Thr	Ser	Asp	Gly	Asp	Ala	Leu
			340						345					350		
	Leu	Leu	Leu													
			355													

201

(234) INFORMATION FOR SEQ ID NO:233:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:233:

10 GGCTTAAGAG CATCATCGTG GTGCTGGTG

29

(235) INFORMATION FOR SEQ ID NO:234:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:234:

20 GTCACCACCA GCACCACGAT GATGCTCTTA AGCC

34

(236) INFORMATION FOR SEQ ID NO:235:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:235:

CAAAGAAAGT ACTGGGCATC GTCTTCTTCC T

31

30 (237) INFORMATION FOR SEQ ID NO:236:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

30

50

50

35

ANSNOOIN, <WD 00221208: 1A>

203

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:240:

5 ACCGCGAGGA TCATGCGTTT CGCCTTCTGC CGCCG

35

(242) INFORMATION FOR SEQ ID NO:241:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs

(B) TYPE: nucleic acid

10 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:241:

GAGACATATT ATCTGCCACG GAGG

24

15 (243) INFORMATION FOR SEQ ID NO:242:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs

(B) TYPE: nucleic acid

20 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:242:

TTGGCATAGA AACCGGACCC AAGG

24

(244) INFORMATION FOR SEQ ID NO:243:

25 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 28 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

30 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:243:

TAAGAATTCC ATAAAAATTA TGGAATGG

28

(245) INFORMATION FOR SEQ ID NO:244:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:244:

CCAGGATCCA GCTGAAGTCT TCCATCATTC

30

(246) INFORMATION FOR SEQ ID NO:245:

- 10 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1071 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:245:

ATGAATGGGG TCTCGGAGGG GACCAGAGGC TGCAGTGACA GGCAACCTGG GGTCCCTGACA 60

CGTGATCGCT CTTGTTCCAG GAAGATGAAC TCTTCCGGAT GCCTGTCTGA GGAGGTGGGG 120

TCCCTCCGCC CACTGACTGT GGTATCCTG TCTGCGTCCA TTGTCGTCGG AGTGCTGGGC 180

AATGGGCTGG TGCTGTGGAT GACTGTCTTC CGTATGGCAC GCACGGTCTC CACCGTCTGC 240

20 TTCTTCCACC TGGCCCTTGC CGATTTTCATG CTCTCACTGT CTCTGCCCAT TGCCATGTAC 300

TATATTGTCT CCAGGCAGTG GCTCCTCGGA GAGTGGGCCT GCAAACCTCTA CATCACCTTT 360

GTGTTCCCTCA GCTACTTTGC CAGTAACTGC CTCCTTGTCT TCATCTCTGT GGACCGTTGC 420

ATCTCTGTCC TCTACCCCGT CTGGGCCCTG AACCACCGCA CTGTGCAGCG GGCAGCTGG 480

CTGGCCTTTG GGGTGTGGCT CCTGGCCGCC GCCTTGTGCT CTGCGCACCT GAAATTCCGG 540

25 ACAACCAGAA AATGGAATGG CTGTACGCAC TGCTACTTGG CGTTCAACTC TGACAATGAG 600

ACTGCCCAGA TTTGGATTGA AGGGGTCGTG GAGGGACACA TTATAGGGAC CATTGGCCAC 660

TTCCTGCTGG GCTTCCTGGG GCCCTTAGCA ATCATAGGCA CCTGCGCCCA CCTCATCCGG 720

GCCAAGCTCT TGCAGGAGGG CTGGGTCCAT GCCAACCGGC CCGCGAGGCT GCTGCTGGTG 780

CTGGTGAGCG CTTTCTTTAT CTTCTGGTCC CCGTTTAACG TGGTGCTGTT GGTCCATCTG 840

30 TGGCGACGGG TGATGCTCAA GGAAATCTAC CACCCCCGGA TGCTGCTCAT CCTCCAGGCT 900

AGCTTTGCCT TGGGCTGTGT CAACAGCAGC CTCAACCCCT TCCTCTACGT CTTGCTGGC 960

205

AGAGATTTC AAGAAAAGTT TTTCCAGTCT TTGACTTCTG CCCTGGCGAG GGCCTTTGGA 1020

GAGGAGGAGT TTCTGTCATC CTGTCCCCGT GGCAACGCCC CCCGGAATG A 1071

(247) INFORMATION FOR SEQ ID NO:246:

- 5 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 356 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

- 10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:246:

Met Asn Gly Val Ser Glu Gly Thr Arg Gly Cys Ser Asp Arg Gln Pro
 1 5 10 15

Gly Val Leu Thr Arg Asp Arg Ser Cys Ser Arg Lys Met Asn Ser Ser
 20 25 30

- 15 Gly Cys Leu Ser Glu Glu Val Gly Ser Leu Arg Pro Leu Thr Val Val
 35 40 45

Ile Leu Ser Ala Ser Ile Val Val Gly Val Leu Gly Asn Gly Leu Val
 50 55 60

- 20 Leu Trp Met Thr Val Phe Arg Met Ala Arg Thr Val Ser Thr Val Cys
 65 70 75 80

Phe Phe His Leu Ala Leu Ala Asp Phe Met Leu Ser Leu Ser Leu Pro
 85 90 95

Ile Ala Met Tyr Tyr Ile Val Ser Arg Gln Trp Leu Leu Gly Glu Trp
 100 105 110

- 25 Ala Cys Lys Leu Tyr Ile Thr Phe Val Phe Leu Ser Tyr Phe Ala Ser
 115 120 125

Asn Cys Leu Leu Val Phe Ile Ser Val Asp Arg Cys Ile Ser Val Leu
 130 135 140

- 30 Tyr Pro Val Trp Ala Leu Asn His Arg Thr Val Gln Arg Ala Ser Trp
 145 150 155 160

Leu Ala Phe Gly Val Trp Leu Leu Ala Ala Ala Leu Cys Ser Ala His
 165 170 175

Leu Lys Phe Arg Thr Thr Arg Lys Trp Asn Gly Cys Thr His Cys Tyr
 180 185 190

- 35 Leu Ala Phe Asn Ser Asp Asn Glu Thr Ala Gln Ile Trp Ile Glu Gly
 195 200 205

Val Val Glu Gly His Ile Ile Gly Thr Ile Gly His Phe Leu Leu Gly
 210 215 220

Phe Leu Gly Pro Leu Ala Ile Ile Gly Thr Cys Ala His Leu Ile Arg
 225 230 235 240

5 Ala Lys Leu Leu Arg Glu Gly Trp Val His Ala Asn Arg Pro Ala Arg
 245 250 255

Leu Leu Leu Val Leu Val Ser Ala Phe Phe Ile Phe Trp Ser Pro Phe
 260 265 270

10 Asn Val Val Leu Leu Val His Leu Trp Arg Arg Val Met Leu Lys Glu
 275 280 285

Ile Tyr His Pro Arg Met Leu Leu Ile Leu Gln Ala Ser Phe Ala Leu
 290 295 300

Gly Cys Val Asn Ser Ser Leu Asn Pro Phe Leu Tyr Val Phe Val Gly
 305 310 315 320

15 Arg Asp Phe Gln Glu Lys Phe Phe Gln Ser Leu Thr Ser Ala Leu Ala
 325 330 335

Arg Ala Phe Gly Glu Glu Glu Phe Leu Ser Ser Cys Pro Arg Gly Asn
 340 345 350

20 Ala Pro Arg Glu
 355

(248) INFORMATION FOR SEQ ID NO:247:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 32 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:247:

GCAGAATTTCG GCGGCCCCAT GGACCTGCCC CC

32

30 (249) INFORMATION FOR SEQ ID NO:248:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:248:

207

GCTGGATCCC CCGAGCAGTG GCGTTACTTC

30

(250) INFORMATION FOR SEQ ID NO:249:

(i) SEQUENCE CHARACTERISTICS:

5

- (A) LENGTH: 903 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:249:

10	ATGGACCTGC CCCC GCAGCT CTCCTTCGGC CTC	TATGTGG CCGCCTTTGC GCTGGGCTTC	60
	CCGCTCAACG TCCTGGCCAT CCGAGGCGCG	ACGGCCCACG CCCGGCTCCG TCTCACCCT	120
	AGCCTGGTCT ACGCCCTGAA CCTGGGCTGC	TCCGACCTGC TGCTGACAGT CTCTCTGCCC	180
	CTGAAGGCGG TGGAGGCGCT AGCCTCCGGG	GCCTGGCCTC TGCCGGCCTC GCTGTGCCCC	240
	GTCTTCGCGG TGGCCCACTT CTTCCCACTC	TATGCCGGCG GGGGCTTCCT GGCCGCCCTG	300
15	AGTGCAGGCC GCTACCTGGG AGCAGCCTTC	CCCTTGGGCT ACCAAGCCTT CCGGAGGCCG	360
	TGCTATTCTT GGGGGGTGTG CGCGGCCATC	TGGGCCCTCG TCCTGTGTCA CCTGGGTCTG	420
	GTCTTTGGGT TGGAGGCTCC AGGAGGCTGG	CTGGACCACA GCAACACCTC CCTGGGCATC	480
	AACACACCGG TCAACGGCTC TCCGGTCTGC	CTGGAGGCCT GGGACCCGGC CTCTGCCGGC	540
	CCGGCCCCGT TCAGCCTCTC TCTCCTGCTC	TTTTTTCTGC CCTTGGCCAT CACAGCCTTC	600
20	TGCTACGTGG GCTGCCTCCG GGC	ACTGGCC CGCTCCGGCC TGACGCACAG	GCGGAAGCTG 660
	CGGGCCGCCT GGGTGGCCGG CGGGGCCCTC	CTCACGCTGC TGCTCTGCGT AGGACCCTAC	720
	AACGCCTCCA ACGTGGCCAG CTTCTGTAC	CCCAATCTAG GAGGCTCCTG GCGGAAGCTG	780
	GGGCTCATCA CGGGTGCCTG GAGTGTGGTG	CTTAATCCGC TGGTGACCGG TTA	TTACTTGGA 840
	AGGGGTCCTG GCCTGAAGAC AGTGTGTGCG	GCAAGAACGC AAGGGGGCAA GTCCCAGAAG	900
25	TAA		903

(251) INFORMATION FOR SEQ ID NO:250:

(i) SEQUENCE CHARACTERISTICS:

30

- (A) LENGTH: 300 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:250:

```

Met Asp Leu Pro Pro Gln Leu Ser Phe Gly Leu Tyr Val Ala Ala Phe
1          5          10          15

5  Ala Leu Gly Phe Pro Leu Asn Val Leu Ala Ile Arg Gly Ala Thr Ala
    20          25          30

His Ala Arg Leu Arg Leu Thr Pro Ser Leu Val Tyr Ala Leu Asn Leu
    35          40          45

10 Gly Cys Ser Asp Leu Leu Leu Thr Val Ser Leu Pro Leu Lys Ala Val
    50          55          60

Glu Ala Leu Ala Ser Gly Ala Trp Pro Leu Pro Ala Ser Leu Cys Pro
65          70          75          80

Val Phe Ala Val Ala His Phe Phe Pro Leu Tyr Ala Gly Gly Gly Phe
    85          90          95

15 Leu Ala Ala Leu Ser Ala Gly Arg Tyr Leu Gly Ala Ala Phe Pro Leu
    100          105          110

Gly Tyr Gln Ala Phe Arg Arg Pro Cys Tyr Ser Trp Gly Val Cys Ala
    115          120          125

20 Ala Ile Trp Ala Leu Val Leu Cys His Leu Gly Leu Val Phe Gly Leu
    130          135          140

Glu Ala Pro Gly Gly Trp Leu Asp His Ser Asn Thr Ser Leu Gly Ile
145          150          155          160

Asn Thr Pro Val Asn Gly Ser Pro Val Cys Leu Glu Ala Trp Asp Pro
    165          170          175

25 Ala Ser Ala Gly Pro Ala Arg Phe Ser Leu Ser Leu Leu Phe Phe
    180          185          190

Leu Pro Leu Ala Ile Thr Ala Phe Cys Tyr Val Gly Cys Leu Arg Ala
    195          200          205

30 Leu Ala Arg Ser Gly Leu Thr His Arg Arg Lys Leu Arg Ala Ala Trp
    210          215          220

Val Ala Gly Gly Ala Leu Leu Thr Leu Leu Leu Cys Val Gly Pro Tyr
225          230          235          240

Asn Ala Ser Asn Val Ala Ser Phe Leu Tyr Pro Asn Leu Gly Gly Ser
    245          250          255

35 Trp Arg Lys Leu Gly Leu Ile Thr Gly Ala Trp Ser Val Val Leu Asn
    260          265          270

Pro Leu Val Thr Gly Tyr Leu Gly Arg Gly Pro Gly Leu Lys Thr Val

```

209

275

280

285

Cys Ala Ala Arg Thr Gln Gly Gly Lys Ser Gln Lys
 290 295 300

(252) INFORMATION FOR SEQ ID NO:251:

- 5 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 31 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- 10 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:251:

CTCAAGCTTA CTCTCTCTCA CCAGTGGCCA C

31

(253) INFORMATION FOR SEQ ID NO:252:

- 15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 24 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

- 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:252:

CCCTCCTCCC CCGGAGGACC TAGC

24

(254) INFORMATION FOR SEQ ID NO:253:

- 25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1041 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:253:

30 ATGGATACAG GCCCCGACCA GTCCTACTTC TCCGGCAATC ACTGGTTCGT CTTCTCGGTG 60
 TACCTTCTCA CTTTCCTGGT GGGGCTCCCC CTCAACCTGC TGGCCCTGGT GGTCTTCGTG 120
 GGCAAGCTGC AGCGCCGCCC GGTGGCCGTG GACGTGCTCC TGCTCAACCT GACCGCCTCG 180
 GACCTGCTCC TGCTGCTGTT CCTGCCTTTC CGCATGGTGG AGGCAGCCAA TGGCATGCAC 240
 TGGCCCCTGC CCTTCATCCT CTGCCCCACTC TCTGGATTCA TCTTCTTCAC CACCATCTAT 300

CTCACCGCCC TCTTCCTGGC AGCTGTGAGC ATTGAACGCT TCCTGAGTGT GGCCACCCCA 360
 CTGTGGTACA AGACCCGGCC GAGGCTGGGG CAGGCAGGTC TGGTGAGTGT GGCCTGCTGG 420
 CTGTTGGCCT CTGCTCACTG CAGCGTGGTC TACGTCATAG AATTCTCAGG GGACATCTCC 480
 CACAGCCAGG GCACCAATGG GACCTGCTAC CTGGAGTTCC GGAAGGACCA GCTAGCCATC 540
 5 CTCCTGCCCCG TCGGGCTGGA GATGGCTGTG GTCCTCTTTG TGGTCCCCT GATCATCACC 600
 AGCTACTGCT ACAGCCGCCT GGTGTGGATC CTCGGCAGAG GGGGCAGCCA CCGCCGGCAG 660
 AGGAGGGTGG CGGGGCTGTT GGCGGCCACG CTGCTCAACT TCCTTGTCTG CTTTGGGCCC 720
 TACAACGTGT CCCATGTCGT GGGCTATATC TCGGTGAAA GCCCGGCATG GAGGATCTAC 780
 GTGACGCTTC TCAGCACCTT GAACTCCTGT GTCGACCCCT TTGTCTACTA CTTCTCCTCC 840
 10 TCCGGGTTCC AAGCCGACTT TCATGAGCTG CTGAGGAGGT TGTGTGGGCT CTGGGGCCAG 900
 TGGCAGCAGG AGAGCAGCAT GGAGCTGAAG GAGCAGAAGG GAGGGGAGGA GCAGAGAGCG 960
 GACCGACCAG CTGAAAGAAA GACCAGTGAA CACTCACAGG GCTGTGGAAC TGGTGGCCAG 1020
 GTGCGCTGTG CTGAAAGCTA G 1041

(255) INFORMATION FOR SEQ ID NO:254:

15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 346 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant
 20 (ii) MOLECULE TYPE: protein
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:254:
 Met Asp Thr Gly Pro Asp Gln Ser Tyr Phe Ser Gly Asn His Trp Phe
 1 5 10 15
 Val Phe Ser Val Tyr Leu Leu Thr Phe Leu Val Gly Leu Pro Leu Asn
 25 20 25 30
 Leu Leu Ala Leu Val Val Phe Val Gly Lys Leu Gln Arg Arg Pro Val
 35 40 45
 Ala Val Asp Val Leu Leu Leu Asn Leu Thr Ala Ser Asp Leu Leu Leu
 50 55 60
 30 Leu Leu Phe Leu Pro Phe Arg Met Val Glu Ala Ala Asn Gly Met His
 65 70 75 80
 Trp Pro Leu Pro Phe Ile Leu Cys Pro Leu Ser Gly Phe Ile Phe Phe
 85 90 95

211

Thr Thr Ile Tyr Leu Thr Ala Leu Phe Leu Ala Ala Val Ser Ile Glu
 100 105 110
 Arg Phe Leu Ser Val Ala His Pro Leu Trp Tyr Lys Thr Arg Pro Arg
 115 120 125
 5 Leu Gly Gln Ala Gly Leu Val Ser Val Ala Cys Trp Leu Leu Ala Ser
 130 135 140
 Ala His Cys Ser Val Val Tyr Val Ile Glu Phe Ser Gly Asp Ile Ser
 145 150 155 160
 10 His Ser Gln Gly Thr Asn Gly Thr Cys Tyr Leu Glu Phe Arg Lys Asp
 165 170 175
 Gln Leu Ala Ile Leu Leu Pro Val Arg Leu Glu Met Ala Val Val Leu
 180 185 190
 Phe Val Val Pro Leu Ile Ile Thr Ser Tyr Cys Tyr Ser Arg Leu Val
 195 200 205
 15 Trp Ile Leu Gly Arg Gly Gly Ser His Arg Arg Gln Arg Arg Val Ala
 210 215 220
 Gly Leu Leu Ala Ala Thr Leu Leu Asn Phe Leu Val Cys Phe Gly Pro
 225 230 235 240
 20 Tyr Asn Val Ser His Val Val Gly Tyr Ile Cys Gly Glu Ser Pro Ala
 245 250 255
 Trp Arg Ile Tyr Val Thr Leu Leu Ser Thr Leu Asn Ser Cys Val Asp
 260 265 270
 Pro Phe Val Tyr Tyr Phe Ser Ser Ser Gly Phe Gln Ala Asp Phe His
 275 280 285
 25 Glu Leu Leu Arg Arg Leu Cys Gly Leu Trp Gly Gln Trp Gln Gln Glu
 290 295 300
 Ser Ser Met Glu Leu Lys Glu Gln Lys Gly Gly Glu Glu Gln Arg Ala
 305 310 315 320
 30 Asp Arg Pro Ala Glu Arg Lys Thr Ser Glu His Ser Gln Gly Cys Gly
 325 330 335
 Thr Gly Gly Gln Val Ala Cys Ala Glu Ser
 340 345

(256) INFORMATION FOR SEQ ID NO:255:

- 35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 31 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:255:

TTTAAGCTTC CCCTCCAGGA TGCTGCCGGA C

31

(257) INFORMATION FOR SEQ ID NO:256:

- 5 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 31 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: not relevant

10 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:256:

GGCGAATTCT GAAGGTCCAG GGAAACTGCT A

31

(258) INFORMATION FOR SEQ ID NO:257:

- 15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 993 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:257:

ATGCTGCCGG ACTGGAAGAG CTCCTTGATC CTCATGGCTT ACATCATCAT CTCCTCACT 60

GGCCTCCCTG CCAACCTCCT GGCCCTGCGG GCCTTTGTGG GGC GGATCCG CCAGCCCCAG 120

CCTGCACCTG TGCACATCCT CCTGCTGAGC CTGACGCTGG CCGACCTCCT CCTGCTGCTG 180

CTGCTGCCCT TCAAGATCAT CGAGGCTGCG TCGAACTTCC GCTGGTACCT GCCCAAGGTC 240

25 GTCTGCGCCC TCACGAGTTT TGGCTTCTAC AGCAGCATCT ACTGCAGCAC GTGGCTCCTG 300

GCGGGCATCA GCATCGAGCG CTACCTGGGA GTGGCTTTCC CCGTGCAGTA CAAGCTCTCC 360

CGCCGGCCTC TGTATGGAGT GATTGCAGCT CTGGTGGCCT GGGTTATGTC CTTTGGTCAC 420

TGCACCATCG TGATCATCGT TCAATACTTG AACACGACTG AGCAGGTCAG AAGTGGCAAT 480

GAAATTACCT GCTACGAGAA CTTACCCGAT AACCAGTTGG ACGTGGTGCT GCCCGTGCGG 540

30 CTGGAGCTGT GCCTGGTGCT CTTCTTCATC CCCATGGCAG TCACCATCTT CTGCTACTGG 600

CGTTTTGTGT GGATCATGCT CTCCAGCCCC CTTGTGGGGG CCCAGAGGCG GCGCCGAGCC 660

GTGGGGCTGG CTGTGGTGAC GCTGCTCAAT TTCCTGGTGT GCTTCGGACC TTACAACGTG 720

213

TCCCACCTGG TGGGGTATCA CCAGAGAAAA AGCCCCTGGT GGCGGTCAAT AGCCGTGGTG 780
 TTCAGTTCAC TCAACGCCAG TCTGGACCCC CTGCTCTTCT ATTTCTCTTC TTCAGTGGTG 840
 CGCAGGGCAT TTGGGAGAGG GCTGCAGGTG CTGCGGAATC AGGGCTCCTC CCTGTTGGGA 900
 CGCAGAGGCA AAGACACAGC AGAGGGGACA AATGAGGACA GGGGTGTGGG TCAAGGAGAA 960
 5 GGGATGCCAA GTTCGGACTT CACTACAGAG TAG 993

(259) INFORMATION FOR SEQ ID NO:258:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 362 amino acids
 (B) TYPE: amino acid
 10 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:258:

15 Met Leu Pro Asp Trp Lys Ser Ser Leu Ile Leu Met Ala Tyr Ile Ile
 1 5 10 15
 Ile Phe Leu Thr Gly Leu Pro Ala Asn Leu Leu Ala Leu Arg Ala Phe
 20 20 25 30
 Val Gly Arg Ile Arg Gln Pro Gln Pro Ala Pro Val His Ile Leu Leu
 35 40 45
 20 Leu Ser Leu Thr Leu Ala Asp Leu Leu Leu Leu Leu Leu Leu Pro Phe
 50 55 60
 Lys Ile Ile Glu Ala Ala Ser Asn Phe Arg Trp Tyr Leu Pro Lys Val
 65 70 75 80
 25 Val Cys Ala Leu Thr Ser Phe Gly Phe Tyr Ser Ser Ile Tyr Cys Ser
 85 90 95
 Thr Trp Leu Leu Ala Gly Ile Ser Ile Glu Arg Tyr Leu Gly Val Ala
 100 105 110
 Phe Pro Val Gln Tyr Lys Leu Ser Arg Arg Pro Leu Tyr Gly Val Ile
 115 120 125
 30 Ala Ala Leu Val Ala Trp Val Met Ser Phe Gly His Cys Thr Ile Val
 130 135 140
 Ile Ile Val Gln Tyr Leu Asn Thr Thr Glu Gln Val Arg Ser Gly Asn
 145 150 155 160
 35 Glu Ile Thr Cys Tyr Glu Asn Phe Thr Asp Asn Gln Leu Asp Val Val
 165 170 175

Leu Pro Val Arg Leu Glu Leu Cys Leu Val Leu Phe Phe Ile Pro Met
 180 185 190
 Ala Val Thr Ile Phe Cys Tyr Trp Arg Phe Val Trp Ile Met Leu Ser
 195 200 205
 5 Gln Pro Leu Val Gly Ala Gln Arg Arg Arg Arg Ala Val Gly Leu Ala
 210 215 220
 Val Val Thr Leu Leu Asn Phe Leu Val Cys Phe Gly Pro Tyr Asn Val
 225 230 235 240
 10 Ser His Leu Val Gly Tyr His Gln Arg Lys Ser Pro Trp Trp Arg Ser
 245 250 255
 Ile Ala Val Val Phe Ser Ser Leu Asn Ala Ser Leu Asp Pro Leu Leu
 260 265 270
 Phe Tyr Phe Ser Ser Ser Val Val Arg Arg Ala Phe Gly Arg Gly Leu
 275 280 285
 15 Gln Val Leu Arg Asn Gln Gly Ser Ser Leu Leu Gly Arg Arg Gly Lys
 290 295 300
 Asp Thr Ala Glu Gly Thr Asn Glu Asp Arg Gly Val Gly Gln Gly Glu
 305 310 315 320
 20 Gly Met Pro Ser Ser Asp Phe Thr Thr Glu
 325 330

(260) INFORMATION FOR SEQ ID NO:259:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: DNA (genomic)
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:259:

CCCAAGCTTC GGGCACCATG GACACCTCCC

30

30 (261) INFORMATION FOR SEQ ID NO:260:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: DNA (genomic)
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:260:

ACAGGATCCA AATGCACAGC ACTGGTAAGC

30

(262) INFORMATION FOR SEQ ID NO:261:

(i) SEQUENCE CHARACTERISTICS:

5

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:261:

10 CTATAACTGG GTTACATGGT TTAAC

25

(263) INFORMATION FOR SEQ ID NO:262:

(i) SEQUENCE CHARACTERISTICS:

15

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:262:

TTTGAATTCA CATATTAATT AGAGACATGG

30

20 (264) INFORMATION FOR SEQ ID NO:263:

(i) SEQUENCE CHARACTERISTICS:

25

- (A) LENGTH: 2724 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:263:

ATGGACACCT CCCGGCTCGG TGTGCTCCTG TCCTTGCCTG TGCTGCTGCA GCTGGCGACC 60

GGGGGCAGCT CTCCCAGGTC TGGTGTGTTG CTGAGGGGCT GCCCCACACA CTGTCATTGC 120

30 GAGCCCGACG GCAGGATGTT GCTCAGGGTG GACTGCTCCG ACCTGGGGCT CTCGGAGCTG 180

CCTTCCAACC TCAGCGTCTT CAGCTCCTAC CTAGACCTCA GTATGAACAA CATCAGTCAG 240

CTGCTCCCGA ATCCCCTGCC CAGTCTCCGC TTCCTGGAGG AGTTACGTCT TCGGGGAAAC 300

GCTCTGACAT ACATTCCCAA GGGAGCATTC ACTGGCCTTT ACAGTCTTAA AGTTCTTATG 360

	CTGCAGAATA ATCAGCTAAG ACACGTACCC ACAGAAGCTC TGCAGAATTT GCGAAGCCTT	420
	CAATCCCTGC GTCTGGATGC TAACCACATC AGCTATGTGC CCCCAGCTG TTTCAGTGGC	480
	CTGCATTCCC TGAGGCACCT GTGGCTGGAT GACAATGCGT TAACAGAAAT CCCCCTCCAG	540
	GCTTTTAGAA GTTTATCGGC ATTGCAAGCC ATGACCTTGG CCCTGAACAA AATACACCAC	600
5	ATACCAGACT ATGCCTTTGG AAACCTCTCC AGCTTGGTAG TTCTACATCT CCATAACAAT	660
	AGAATCCACT CCCTGGGAAA GAAATGCTTT GATGGGCTCC ACAGCCTAGA GACTTTAGAT	720
	TTAAATTACA ATAACCTTGA TGAATTCCCC ACTGCAATTA GGACACTCTC CAACCTTAAA	780
	GAACTAGGAT TTCATAGCAA CAATATCAGG TCGATACCTG AGAAAGCATT TGTAGGCAAC	840
	CCTTCTCTTA TTACAATACA TTTCTATGAC AATCCCATCC AATTTGTTGG GAGATCTGCT	900
10	TTTCAACATT TACCTGAACT AAGAACACTG ACTCTGAATG GTGCCTCACA AATAACTGAA	960
	TTTCCTGATT TAACTGGAAC TGCAAACCTG GAGAGTCTGA CTTTAACTGG AGCACAGATC	1020
	TCATCTCTTC CTCAAACCGT CTGCAATCAG TTACCTAATC TCCAAGTGCT AGATCTGTCT	1080
	TACAACCTAT TAGAAGATTT ACCCAGTTTT TCAGTCTGCC AAAAGCTTCA GAAATTGAC	1140
	CTAAGACATA ATGAAATCTA CGAAATTAAA GTTGACACTT TCCAGCAGTT GCTTAGCCTC	1200
15	CGATCGCTGA ATTTGGCTTG GAACAAAATT GCTATTATTC ACCCAATGC ATTTTCCACT	1260
	TTGCCATCCC TAATAAAGCT GGACCTATCG TCCAACCTCC TGTCGTCTTT TCCTATAACT	1320
	GGGTACATG GTTTAACTCA CTTAAAATTA ACAGGAAATC ATGCCTTACA GAGCTTGATA	1380
	TCATCTGAAA ACTTTCCAGA ACTCAAGGTT ATAGAAATGC CTTATGCTTA CCAGTGCTGT	1440
	GCATTGGAG TGTGTGAGAA TGCCTATAAG ATTTCTAATC AATGGAATAA AGGTGACAAC	1500
20	AGCAGTATGG ACGACCTTCA TAAGAAAGAT GCTGGAATGT TTCAGGCTCA AGATGAACGT	1560
	GACCTTGAAG ATTTCTGCT TGAATTTGAG GAAGACCTGA AAGCCCTTCA TTCACTGCAG	1620
	TGTTACCTT CCCCAGGCCC CTTCAAACCC TGTGAACACC TGCTTGATGG CTGGCTGATC	1680
	AGAATTGGAG TGTGGACCAT AGCAGTTCTG GCACTTACTT GTAATGCTTT GGTGACTTCA	1740
	ACAGTTTTCA GATCCCCCTCT GTACATTTC CCCATTAAAC TGTTAATTGG GGTCATCGCA	1800
25	GCAGTGAACA TGCTCACGGG AGTCTCCAGT GCCGTGCTGG CTGGTGTGGA TCGTTTCACT	1860
	TTTGGCAGCT TTGCACGACA TGGTGCTGG TGGGAGAATG GGGTTGGTTG CCATGTCATT	1920
	GGTTTTTTGT CCATTTTTC TTCAGAATCA TCTGTTTTCC TGCTTACTCT GGCAGCCCTG	1980
	GAGCGTGGGT TCTCTGTGAA ATATTCTGCA AAATTTGAAA CGAAAGCTCC ATTTTCTAGC	2040

217

CTGAAAGTAA TCATTTTGCT CTGTGCCCTG CTGGCCTTGA CCATGGCCGC AGTTCCCCTG 2100
 CTGGGTGGCA GCAAGTATGG CGCCTCCCCT CTCTGCCTGC CTTTGCCTTT TGGGGAGCCC 2160
 AGCACCATGG GCTACATGGT CGCTCTCATC TTGCTCAATT CCCTTTGCTT CCTCATGATG 2220
 ACCATTGCCT ACACCAAGCT CTA CTGCAAT TTGGACAAGG GAGACCTGGA GAATATTTGG 2280
 5 GACTGCTCTA TGGTAAACA CATTGCCCTG TTGCTCTTCA CCAACTGCAT CCTAAACTGC 2340
 CCTGTGGCTT TCTTGTCTT CTCCTCTTTA ATAAACCTTA CATTTATCAG TCCTGAAGTA 2400
 ATTAAGTTTA TCCTTCTGGT GGTAGTCCCA CTTCTGTCAT GTCTCAATCC CTTTCTCTAC 2460
 ATCTTGTTCA ATCCTCACTT TAAGGAGGAT CTGGTGAGCC TGAGAAAGCA AACCTACGTC 2520
 TGGACAAGAT CAAAACACCC AAGCTTGATG TCAATTA ACT CTGATGATGT CGAAAAACAG 2580
 10 TCCTGTGACT CAACTCAAGC CTTGGTAACC TTTACCAGCT CCAGCATCAC TTATGACCTG 2640
 CCTCCCAGTT CCGTGCCATC ACCAGCTTAT CCAGTGACTG AGAGCTGCCA TCTTTCCTCT 2700
 GTGGCATTG TCCCATGTCT CTAA 2724

(265) INFORMATION FOR SEQ ID NO:264:

- 15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 907 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

- 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:264:

Met Asp Thr Ser Arg Leu Gly Val Leu Leu Ser Leu Pro Val Leu Leu
 1 5 10 15
 Gln Leu Ala Thr Gly Gly Ser Ser Pro Arg Ser Gly Val Leu Leu Arg
 20 25 30
 25 Gly Cys Pro Thr His Cys His Cys Glu Pro Asp Gly Arg Met Leu Leu
 35 40 45
 Arg Val Asp Cys Ser Asp Leu Gly Leu Ser Glu Leu Pro Ser Asn Leu
 50 55 60
 30 Ser Val Phe Thr Ser Tyr Leu Asp Leu Ser Met Asn Asn Ile Ser Gln
 65 70 75 80
 Leu Leu Pro Asn Pro Leu Pro Ser Leu Arg Phe Leu Glu Glu Leu Arg
 85 90 95
 Leu Ala Gly Asn Ala Leu Thr Tyr Ile Pro Lys Gly Ala Phe Thr Gly

	100	105	110
	Leu Tyr Ser Leu Lys Val	Leu Met Leu Gln Asn Asn	Gln Leu Arg His
	115	120	125
5	Val Pro Thr Glu Ala Leu	Gln Asn Leu Arg Ser	Leu Gln Ser Leu Arg
	130	135	140
	Leu Asp Ala Asn His Ile	Ser Tyr Val Pro Pro	Ser Cys Phe Ser Gly
	145	150	155 160
	Leu His Ser Leu Arg His	Leu Trp Leu Asp Asp	Asn Ala Leu Thr Glu
	165	170	175
10	Ile Pro Val Gln Ala Phe	Arg Ser Leu Ser Ala	Leu Gln Ala Met Thr
	180	185	190
	Leu Ala Leu Asn Lys Ile	His His Ile Pro Asp	Tyr Ala Phe Gly Asn
	195	200	205
15	Leu Ser Ser Leu Val Val	Leu His Leu His Asn	Asn Arg Ile His Ser
	210	215	220
	Leu Gly Lys Lys Cys Phe	Asp Gly Leu His Ser	Leu Glu Thr Leu Asp
	225	230	235 240
	Leu Asn Tyr Asn Asn Leu	Asp Glu Phe Pro Thr	Ala Ile Arg Thr Leu
	245	250	255
20	Ser Asn Leu Lys Glu Leu	Gly Phe His Ser Asn	Asn Ile Arg Ser Ile
	260	265	270
	Pro Glu Lys Ala Phe Val	Gly Asn Pro Ser Leu	Ile Thr Ile His Phe
	275	280	285
25	Tyr Asp Asn Pro Ile Gln	Phe Val Gly Arg Ser	Ala Phe Gln His Leu
	290	295	300
	Pro Glu Leu Arg Thr Leu	Thr Leu Asn Gly Ala	Ser Gln Ile Thr Glu
	305	310	315 320
	Phe Pro Asp Leu Thr Gly	Thr Ala Asn Leu Glu	Ser Leu Thr Leu Thr
	325	330	335
30	Gly Ala Gln Ile Ser Ser	Leu Pro Gln Thr Val	Cys Asn Gln Leu Pro
	340	345	350
	Asn Leu Gln Val Leu Asp	Leu Ser Tyr Asn Leu	Leu Glu Asp Leu Pro
	355	360	365
35	Ser Phe Ser Val Cys Gln	Lys Leu Gln Lys Ile	Asp Leu Arg His Asn
	370	375	380
	Glu Ile Tyr Glu Ile Lys	Val Asp Thr Phe Gln	Gln Leu Leu Ser Leu
	385	390	395 400

219

Arg Ser Leu Asn Leu Ala Trp Asn Lys Ile Ala Ile Ile His Pro Asn
 405 410 415

Ala Phe Ser Thr Leu Pro Ser Leu Ile Lys Leu Asp Leu Ser Ser Asn
 420 425 430

5 Leu Leu Ser Ser Phe Pro Ile Thr Gly Leu His Gly Leu Thr His Leu
 435 440 445

Lys Leu Thr Gly Asn His Ala Leu Gln Ser Leu Ile Ser Ser Glu Asn
 450 455 460

10 Phe Pro Glu Leu Lys Val Ile Glu Met Pro Tyr Ala Tyr Gln Cys Cys
 465 470 475 480

Ala Phe Gly Val Cys Glu Asn Ala Tyr Lys Ile Ser Asn Gln Trp Asn
 485 490 495

Lys Gly Asp Asn Ser Ser Met Asp Asp Leu His Lys Lys Asp Ala Gly
 500 505 510

15 Met Phe Gln Ala Gln Asp Glu Arg Asp Leu Glu Asp Phe Leu Leu Asp
 515 520 525

Phe Glu Glu Asp Leu Lys Ala Leu His Ser Val Gln Cys Ser Pro Ser
 530 535 540

20 Pro Gly Pro Phe Lys Pro Cys Glu His Leu Leu Asp Gly Trp Leu Ile
 545 550 555 560

Arg Ile Gly Val Trp Thr Ile Ala Val Leu Ala Leu Thr Cys Asn Ala
 565 570 575

Leu Val Thr Ser Thr Val Phe Arg Ser Pro Leu Tyr Ile Ser Pro Ile
 580 585 590

25 Lys Leu Leu Ile Gly Val Ile Ala Ala Val Asn Met Leu Thr Gly Val
 595 600 605

Ser Ser Ala Val Leu Ala Gly Val Asp Ala Phe Thr Phe Gly Ser Phe
 610 615 620

30 Ala Arg His Gly Ala Trp Trp Glu Asn Gly Val Gly Cys His Val Ile
 625 630 635 640

Gly Phe Leu Ser Ile Phe Ala Ser Glu Ser Ser Val Phe Leu Leu Thr
 645 650 655

Leu Ala Ala Leu Glu Arg Gly Phe Ser Val Lys Tyr Ser Ala Lys Phe
 660 665 670

35 Glu Thr Lys Ala Pro Phe Ser Ser Leu Lys Val Ile Ile Leu Leu Cys
 675 680 685

Ala Leu Leu Ala Leu Thr Met Ala Ala Val Pro Leu Leu Gly Gly Ser

	690		695		700
	Lys Tyr Gly Ala Ser Pro Leu Cys Leu Pro Leu Pro Phe Gly Glu Pro				
	705		710		715 720
5	Ser Thr Met Gly Tyr Met Val Ala Leu Ile Leu Leu Asn Ser Leu Cys				
		725		730	735
	Phe Leu Met Met Thr Ile Ala Tyr Thr Lys Leu Tyr Cys Asn Leu Asp				
		740		745	750
	Lys Gly Asp Leu Glu Asn Ile Trp Asp Cys Ser Met Val Lys His Ile				
		755		760	765
10	Ala Leu Leu Leu Phe Thr Asn Cys Ile Leu Asn Cys Pro Val Ala Phe				
		770		775	780
	Leu Ser Phe Ser Ser Leu Ile Asn Leu Thr Phe Ile Ser Pro Glu Val				
		785		790	795 800
15	Ile Lys Phe Ile Leu Leu Val Val Val Pro Leu Pro Ala Cys Leu Asn				
		805		810	815
	Pro Leu Leu Tyr Ile Leu Phe Asn Pro His Phe Lys Glu Asp Leu Val				
		820		825	830
	Ser Leu Arg Lys Gln Thr Tyr Val Trp Thr Arg Ser Lys His Pro Ser				
		835		840	845
20	Leu Met Ser Ile Asn Ser Asp Asp Val Glu Lys Gln Ser Cys Asp Ser				
		850		855	860
	Thr Gln Ala Leu Val Thr Phe Thr Ser Ser Ser Ile Thr Tyr Asp Leu				
		865		870	875 880
25	Pro Pro Ser Ser Val Pro Ser Pro Ala Tyr Pro Val Thr Glu Ser Cys				
		885		890	895
	His Leu Ser Ser Val Ala Phe Val Pro Cys Leu				
		900		905	

(266) INFORMATION FOR SEQ ID NO:265:

- 30 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- 35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:265:

CGGAAGCTGC GGGCCAAATG GGTGGCCGCGC

30

221

(267) INFORMATION FOR SEQ ID NO:266:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:266:

CAGAGGAGGG TGAAGGGGCT GTTGGCG

27

10 (268) INFORMATION FOR SEQ ID NO:267:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:267:

GGCGGCGCCG AGCCAAGGGG CTGGCTGTGG

30

(269) INFORMATION FOR SEQ ID NO:268:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 32 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:268:

GGGACTGCTC TATGAAAAA CACATTGCCC TG

32

(270) INFORMATION FOR SEQ ID NO:269:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1071 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:269:

ATGAATGGGG TCTCGGAGGG GACCAGAGGC TGCAGTGACA GGCAACCTGG GGTCTTGACA

60

CGTGATCGCT CTTGTTCCAG GAAGATGAAC TCTTCCGGAT GCCTGTCTGA GGAGGTGGGG 120
 TCCCTCCGCC CACTGACTGT GGTATCCTG TCTGCGTCCA TTGTCGTCGG AGTGCTGGGC 180
 AATGGGCTGG TGCTGTGGAT GACTGTCTTC CGTATGGCAC GCACGGTCTC CACCGTCTGC 240
 TTCTTCCACC TGGCCCTTGC CGATTTCATG CTCTCACTGT CTCTGCCCAT TGCCATGTAC 300
 5 TATATTGTCT CCAGGCAGTG GCTCCTCGGA GAGTGGGCCT GCAAACCTA CATCACCTTT 360
 GTGTTCTCA GCTACTTTGC CAGTAACTGC CTCCTTGTCT TCATCTCTGT GGACCGTTGC 420
 ATCTCTGTCC TCTACCCCGT CTGGGCCCTG AACCACCGCA CTGTGCAGCG GGCGAGCTGG 480
 CTGGCCTTTG GGGTGTGGCT CCTGGCCGCC GCCTTGTGCT CTGCGCACCT GAAATTCCGG 540
 ACAACCAGAA AATGGAATGG CTGTACGCAC TGCTACTTGG CGTTCAACTC TGACAATGAG 600
 10 ACTGCCCAGA TTTGGATTGA AGGGGTCGTG GAGGGACACA TTATAGGGAC CATTGGCCAC 660
 TTCCTGCTGG GCTTCCTGGG GCCCTTAGCA ATCATAGGCA CCTGCGCCCA CCTCATCCGG 720
 GCCAAGCTCT TGC GGAGGG CTGGGTCCAT GCCAACCGGC CCAAGAGGCT GCTGCTGGTG 780
 CTGGTGAGCG CTTTCTTTAT CTTCTGGTCC CCGTTTAACG TGGTGCTGTT GGTCCATCTG 840
 TGGCGACGGG TGATGCTCAA GGAAATCTAC CACCCCCGGA TGCTGCTCAT CCTCCAGGCT 900
 15 AGCTTTGCCT TGGGCTGTGT CAACAGCAGC CTCAACCCT TCCTCTACGT CTTGTTGGC 960
 AGAGATTTC AAGAAAAGTT TTTCCAGTCT TTGACTTCTG CCCTGGCGAG GCGGTTTGA 1020
 GAGGAGGAGT TTCTGTCATC CTGTCCCCGT GGCAACGCCC CCCGGAATG A 1071

(271) INFORMATION FOR SEQ ID NO:270:

- 20 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 356 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

- 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:270:

Met Asn Gly Val Ser Glu Gly Thr Arg Gly Cys Ser Asp Arg Gln Pro
 1 5 10 15

Gly Val Leu Thr Arg Asp Arg Ser Cys Ser Arg Lys Met Asn Ser Ser
 20 25 30

- 30 Gly Cys Leu Ser Glu Glu Val Gly Ser Leu Arg Pro Leu Thr Val Val
 35 40 45

223

	Ile	Leu	Ser	Ala	Ser	Ile	Val	Val	Gly	Val	Leu	Gly	Asn	Gly	Leu	Val	
	50						55					60					
	Leu	Trp	Met	Thr	Val	Phe	Arg	Met	Ala	Arg	Thr	Val	Ser	Thr	Val	Cys	
	65					70					75					80	
5	Phe	Phe	His	Leu	Ala	Leu	Ala	Asp	Phe	Met	Leu	Ser	Leu	Ser	Leu	Pro	
				85						90					95		
	Ile	Ala	Met	Tyr	Tyr	Ile	Val	Ser	Arg	Gln	Trp	Leu	Leu	Gly	Glu	Trp	
				100					105					110			
10	Ala	Cys	Lys	Leu	Tyr	Ile	Thr	Phe	Val	Phe	Leu	Ser	Tyr	Phe	Ala	Ser	
			115					120					125				
	Asn	Cys	Leu	Leu	Val	Phe	Ile	Ser	Val	Asp	Arg	Cys	Ile	Ser	Val	Leu	
			130				135					140					
	Tyr	Pro	Val	Trp	Ala	Leu	Asn	His	Arg	Thr	Val	Gln	Arg	Ala	Ser	Trp	
	145					150					155					160	
15	Leu	Ala	Phe	Gly	Val	Trp	Leu	Leu	Ala	Ala	Ala	Leu	Cys	Ser	Ala	His	
				165					170						175		
	Leu	Lys	Phe	Arg	Thr	Thr	Arg	Lys	Trp	Asn	Gly	Cys	Thr	His	Cys	Tyr	
				180					185					190			
20	Leu	Ala	Phe	Asn	Ser	Asp	Asn	Glu	Thr	Ala	Gln	Ile	Trp	Ile	Glu	Gly	
			195					200					205				
	Val	Val	Glu	Gly	His	Ile	Ile	Gly	Thr	Ile	Gly	His	Phe	Leu	Leu	Gly	
			210				215					220					
	Phe	Leu	Gly	Pro	Leu	Ala	Ile	Ile	Gly	Thr	Cys	Ala	His	Leu	Ile	Arg	
	225					230					235					240	
25	Ala	Lys	Leu	Leu	Arg	Glu	Gly	Trp	Val	His	Ala	Asn	Arg	Pro	Lys	Arg	
					245					250					255		
	Leu	Leu	Leu	Val	Leu	Val	Ser	Ala	Phe	Phe	Ile	Phe	Trp	Ser	Pro	Phe	
				260					265					270			
30	Asn	Val	Val	Leu	Leu	Val	His	Leu	Trp	Arg	Arg	Val	Met	Leu	Lys	Glu	
			275					280					285				
	Ile	Tyr	His	Pro	Arg	Met	Leu	Leu	Ile	Leu	Gln	Ala	Ser	Phe	Ala	Leu	
			290				295					300					
	Gly	Cys	Val	Asn	Ser	Ser	Leu	Asn	Pro	Phe	Leu	Tyr	Val	Phe	Val	Gly	
	305					310					315					320	
35	Arg	Asp	Phe	Gln	Glu	Lys	Phe	Phe	Gln	Ser	Leu	Thr	Ser	Ala	Leu	Ala	
				325						330					335		
	Arg	Ala	Phe	Gly	Glu	Glu	Glu	Phe	Leu	Ser	Ser	Cys	Pro	Arg	Gly	Asn	

340

345

350

Ala Pro Arg Glu
355

(272) INFORMATION FOR SEQ ID NO:271:

- 5 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 903 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:271:

ATGGACCTGC CCCCCGAGCT CTCCTTCGGC CTCTATGTGG CCGCCTTTGC GCTGGGCTTC 60
 CCGCTCAACG TCCTGGCCAT CCGAGGCGCG ACGGCCACG CCCGGCTCCG TCTACCCCT 120
 15 AGCCTGGTCT ACGCCCTGAA CCTGGGCTGC TCCGACCTGC TGCTGACAGT CTCTCTGCCC 180
 CTGAAGGCGG TGGAGGCGCT AGCCTCCGGG GCCTGGCCTC TGCCGGCCTC GCTGTGCCCC 240
 GTCTTCGCGG TGGCCCACTT CTTCCCACTC TATGCCGGCG GGGGCTTCCT GGCCGCCCTG 300
 AGTGCAGGCC GCTACCTGGG AGCAGCCTTC CCCTTGGGCT ACCAAGCCTT CCGGAGGCCG 360
 TGCTATTCTT GGGGGGTGTG CGCGGCCATC TGGGCCCTCG TCCTGTGTCA CCTGGGTCTG 420
 20 GTCTTTGGGT TGGAGGCTCC AGGAGGCTGG CTGGACCACA GCAACACCTC CCTGGGCATC 480
 AACACACCGG TCAACGGCTC TCCGGTCTGC CTGGAGGCCT GGGACCCGGC CTCTGCCGGC 540
 CCGGCCCCGT TCAGCCTCTC TCTCCTGCTC TTTTCTCTGC CCTTGGCCAT CACAGCCTTC 600
 TGCTACGTGG GCTGCCTCCG GGCCTGGCC CGCTCCGGCC TGACGCACAG GCGGAAGCTG 660
 CGGGCCAAAT GGGTGGCCGG CGGGGCCCTC CTCACGCTGC TGCTCTGCGT AGGACCCTAC 720
 25 AACGCCTCCA ACGTGGCCAG CTTCTGTAC CCCAATCTAG GAGGCTCCTG GCGGAAGCTG 780
 GGGCTCATCA CGGGTGCTG GAGTGTGGTG CTTAATCCGC TGGTGACCGG TTACTTGGGA 840
 AGGGGTCCTG GCCTGAAGAC AGTGTGTGCG GCAAGAACGC AAGGGGGCAA GTCCCAGAAG 900
 TAA 903

(273) INFORMATION FOR SEQ ID NO:272:

- 30 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 300 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:

225

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:272:

5	Met	Asp	Leu	Pro	Pro	Gln	Leu	Ser	Phe	Gly	Leu	Tyr	Val	Ala	Ala	Phe	1	5	10	15
	Ala	Leu	Gly	Phe	Pro	Leu	Asn	Val	Leu	Ala	Ile	Arg	Gly	Ala	Thr	Ala	20	25	30	
	His	Ala	Arg	Leu	Arg	Leu	Thr	Pro	Ser	Leu	Val	Tyr	Ala	Leu	Asn	Leu	35	40	45	
10	Gly	Cys	Ser	Asp	Leu	Leu	Leu	Thr	Val	Ser	Leu	Pro	Leu	Lys	Ala	Val	50	55	60	
	Glu	Ala	Leu	Ala	Ser	Gly	Ala	Trp	Pro	Leu	Pro	Ala	Ser	Leu	Cys	Pro	65	70	75	
15	Val	Phe	Ala	Val	Ala	His	Phe	Phe	Pro	Leu	Tyr	Ala	Gly	Gly	Gly	Phe	85	90	95	
	Leu	Ala	Ala	Leu	Ser	Ala	Gly	Arg	Tyr	Leu	Gly	Ala	Ala	Phe	Pro	Leu	100	105	110	
	Gly	Tyr	Gln	Ala	Phe	Arg	Arg	Pro	Cys	Tyr	Ser	Trp	Gly	Val	Cys	Ala	115	120	125	
20	Ala	Ile	Trp	Ala	Leu	Val	Leu	Cys	His	Leu	Gly	Leu	Val	Phe	Gly	Leu	130	135	140	
	Glu	Ala	Pro	Gly	Gly	Trp	Leu	Asp	His	Ser	Asn	Thr	Ser	Leu	Gly	Ile	145	150	155	
25	Asn	Thr	Pro	Val	Asn	Gly	Ser	Pro	Val	Cys	Leu	Glu	Ala	Trp	Asp	Pro	165	170	175	
	Ala	Ser	Ala	Gly	Pro	Ala	Arg	Phe	Ser	Leu	Ser	Leu	Leu	Leu	Phe	Phe	180	185	190	
	Leu	Pro	Leu	Ala	Ile	Thr	Ala	Phe	Cys	Tyr	Val	Gly	Cys	Leu	Arg	Ala	195	200	205	
30	Leu	Ala	Arg	Ser	Gly	Leu	Thr	His	Arg	Arg	Lys	Leu	Arg	Ala	Lys	Trp	210	215	220	
	Val	Ala	Gly	Gly	Ala	Leu	Leu	Thr	Leu	Leu	Leu	Cys	Val	Gly	Pro	Tyr	225	230	235	
35	Asn	Ala	Ser	Asn	Val	Ala	Ser	Phe	Leu	Tyr	Pro	Asn	Leu	Gly	Gly	Ser	245	250	255	
	Trp	Arg	Lys	Leu	Gly	Leu	Ile	Thr	Gly	Ala	Trp	Ser	Val	Val	Leu	Asn				

260

265

270

Pro Leu Val Thr Gly Tyr Leu Gly Arg Gly Pro Gly Leu Lys Thr Val
 275 280 285

5 Cys Ala Ala Arg Thr Gln Gly Gly Lys Ser Gln Lys
 290 295 300

(274) INFORMATION FOR SEQ ID NO:273:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1041 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:273:

10 ATGGATACAG GCCCGACCA GTCCTACTTC TCCGGCAATC ACTGGTTCGT CTTCTCGGTG 60
 15 TACCTTCTCA CTTTCCTGGT GGGGCTCCCC CTCAACCTGC TGGCCCTGGT GGTCTTCGTG 120
 GGCAAGCTGC AGCGCCGCCC GGTGGCCGTG GACGTGCTCC TGCTCAACCT GACCGCCTCG 180
 GACCTGCTCC TGCTGCTGTT CCTGCCTTTC CGCATGGTGG AGGCAGCCAA TGGCATGCAC 240
 TGGCCCCTGC CCTTCATCCT CTGCCCCTC TCTGGATTCA TCTTCTTCAC CACCATCTAT 300
 CTCACCGCCC TCTTCCTGGC AGCTGTGAGC ATTGAACGCT TCCTGAGTGT GGCCACCCA 360
 20 CTGTGGTACA AGACCCGGCC GAGGCTGGGG CAGGCAGGTC TGGTGAGTGT GGCCTGCTGG 420
 CTGTTGGCCT CTGCTCACTG CAGCGTGGTC TACGTCATAG AATTCTCAGG GGACATCTCC 480
 CACAGCCAGG GCACCAATGG GACCTGCTAC CTGGAGTTCC GGAAGGACCA GCTAGCCATC 540
 CTCCTGCCCC TCGCGCTGGA GATGGCTGTG GTCCTCTTTG TGGTCCCGCT GATCATCACC 600
 AGCTACTGCT ACAGCCGCCT GGTGTGGATC CTCGGCAGAG GGGGCAGCCA CCGCCGGCAG 660
 25 AGGAGGGTGA AGGGGCTGTT GGCGGCCACG CTGCTCAACT TCCTTGCTG CTTTGGGCCC 720
 TACAACGTGT CCCATGTCGT GGGCTATATC TGCGGTGAAA GCGCGGCATG GAGGATCTAC 780
 GTGACGCTTC TCAGCACCCCT GAACTCCTGT GTCGACCCCT TTGTCTACTA CTTCTCCTCC 840
 TCCGGGTTC AAGCCGACTT TCATGAGCTG CTGAGGAGGT TGTGTGGGCT CTGGGGCCAG 900
 TGGCAGCAGG AGAGCAGCAT GGAGCTGAAG GAGCAGAAGG GAGGGGAGGA GCAGAGAGCG 960
 30 GACCGACCAG CTGAAAGAAA GACCAGTGAA CACTCACAGG GCTGTGGAAC TGGTGGCCAG 1020
 GTGGCCTGTG CTGAAAGCTA G 1041

227

(275) INFORMATION FOR SEQ ID NO:274:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 346 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:274:

Met Asp Thr Gly Pro Asp Gln Ser Tyr Phe Ser Gly Asn His Trp Phe
 1 5 10 15

Val Phe Ser Val Tyr Leu Leu Thr Phe Leu Val Gly Leu Pro Leu Asn
 20 25 30

Leu Leu Ala Leu Val Val Phe Val Gly Lys Leu Gln Arg Arg Pro Val
 35 40 45

Ala Val Asp Val Leu Leu Leu Asn Leu Thr Ala Ser Asp Leu Leu Leu
 50 55 60

Leu Leu Phe Leu Pro Phe Arg Met Val Glu Ala Ala Asn Gly Met His
 65 70 75 80

Trp Pro Leu Pro Phe Ile Leu Cys Pro Leu Ser Gly Phe Ile Phe Phe
 85 90 95

Thr Thr Ile Tyr Leu Thr Ala Leu Phe Leu Ala Ala Val Ser Ile Glu
 100 105 110

Arg Phe Leu Ser Val Ala His Pro Leu Trp Tyr Lys Thr Arg Pro Arg
 115 120 125

Leu Gly Gln Ala Gly Leu Val Ser Val Ala Cys Trp Leu Leu Ala Ser
 130 135 140

Ala His Cys Ser Val Val Tyr Val Ile Glu Phe Ser Gly Asp Ile Ser
 145 150 155 160

His Ser Gln Gly Thr Asn Gly Thr Cys Tyr Leu Glu Phe Arg Lys Asp
 165 170 175

Gln Leu Ala Ile Leu Leu Pro Val Arg Leu Glu Met Ala Val Val Leu
 180 185 190

Phe Val Val Pro Leu Ile Ile Thr Ser Tyr Cys Tyr Ser Arg Leu Val
 195 200 205

Trp Ile Leu Gly Arg Gly Gly Ser His Arg Arg Gln Arg Arg Val Lys
 210 215 220

Gly Leu Leu Ala Ala Thr Leu Leu Asn Phe Leu Val Cys Phe Gly Pro

225 230 235 240
 Tyr Asn Val Ser His Val Val Gly Tyr Ile Cys Gly Glu Ser Pro Ala
 245 250 255
 5 Trp Arg Ile Tyr Val Thr Leu Leu Ser Thr Leu Asn Ser Cys Val Asp
 260 265 270
 Pro Phe Val Tyr Tyr Phe Ser Ser Ser Gly Phe Gln Ala Asp Phe His
 275 280 285
 Glu Leu Leu Arg Arg Leu Cys Gly Leu Trp Gly Gln Trp Gln Gln Glu
 290 295 300
 10 Ser Ser Met Glu Leu Lys Glu Gln Lys Gly Gly Glu Glu Gln Arg Ala
 305 310 315 320
 Asp Arg Pro Ala Glu Arg Lys Thr Ser Glu His Ser Gln Gly Cys Gly
 325 330 335
 15 Thr Gly Gly Gln Val Ala Cys Ala Glu Ser
 340 345

(276) INFORMATION FOR SEQ ID NO:275:

20 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 993 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:275:

ATGCTGCCGG ACTGGAAGAG CTCCTTGATC CTCATGGCTT ACATCATCAT CTCCTCACT 60
 25 GGCCTCCCTG CCAACCTCCT GGCCCTGCGG GCCTTTGTGG GCGGATCCG CCAGCCCCAG 120
 CCTGCACCTG TGCACATCCT CCTGCTGAGC CTGACGCTGG CCGACCTCCT CCTGCTGCTG 180
 CTGCTGCCCT TCAAGATCAT CGAGGCTGCG TCGAACTTCC GCTGGTACCT GCCCAAGGTC 240
 GTCTGCGCCC TCACGAGTTT TGGCTTCTAC AGCAGCATCT ACTGCAGCAC GTGGCTCCTG 300
 GCGGGCATCA GCATCGAGCG CTACCTGGGA GTGGCTTTCC CCGTGCAGTA CAAGCTCTCC 360
 30 CGCCGGCCTC TGTATGGAGT GATTGCAGCT CTGGTGGCCT GGGTTATGTC CTTTGGTCAC 420
 TGCACCATCG TGATCATCGT TCAATACTTG AACACGACTG AGCAGGTCAG AAGTGGCAAT 480
 GAAATTACCT GCTACGAGAA CTTACCGAT AACCAGTTGG ACGTGGTGCT GCCCGTGCGG 540
 CTGGAGCTGT GCCTGGTGCT CTTCTTCATC CCCATGGCAG TCACCATCTT CTGCTACTGG 600

229

CGTTTTGTGT GGATCATGCT CTCCCAGCCC CTTGTGGGGG CCCAGAGGCG GCGCCGAGCC 660
 AAGGGGCTGG CTGTGGTGAC GCTGCTCAAT TTCCTGGTGT GCTTCGGACC TTACAACGTG 720
 TCCCACCTGG TGGGGTATCA CCAGAGAAAA AGCCCCTGGT GGCGGTCAAT AGCCGTGGTG 780
 TTCAGTTCAC TCAACGCCAG TCTGGACCCC CTGCTCTTCT ATTTCTCTTC TTCAGTGGTG 840
 5 CGCAGGGCAT TTGGGAGAGG GCTGCAGGTG CTGCGGAATC AGGGCTCCTC CCTGTTGGGA 900
 CGCAGAGGCA AAGACACAGC AGAGGGGACA AATGAGGACA GGGGTGTGGG TCAAGGAGAA 960
 GGGATGCCAA GTTCGGACTT CACTACAGAG TAG 993

(277) INFORMATION FOR SEQ ID NO:276:

10 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 330 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant
 (ii) MOLECULE TYPE: protein

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:276:

Met Leu Pro Asp Trp Lys Ser Ser Leu Ile Leu Met Ala Tyr Ile Ile
 1 5 10 15
 Ile Phe Leu Thr Gly Leu Pro Ala Asn Leu Leu Ala Leu Arg Ala Phe
 20 25 30
 20 Val Gly Arg Ile Arg Gln Pro Gln Pro Ala Pro Val His Ile Leu Leu
 35 40 45
 Leu Ser Leu Thr Leu Ala Asp Leu Leu Leu Leu Leu Leu Pro Phe
 50 55 60
 25 Lys Ile Ile Glu Ala Ala Ser Asn Phe Arg Trp Tyr Leu Pro Lys Val
 65 70 75 80
 Val Cys Ala Leu Thr Ser Phe Gly Phe Tyr Ser Ser Ile Tyr Cys Ser
 85 90 95
 Thr Trp Leu Leu Ala Gly Ile Ser Ile Glu Arg Tyr Leu Gly Val Ala
 100 105 110
 30 Phe Pro Val Gln Tyr Lys Leu Ser Arg Arg Pro Leu Tyr Gly Val Ile
 115 120 125
 Ala Ala Leu Val Ala Trp Val Met Ser Phe Gly His Cys Thr Ile Val
 130 135 140
 35 Ile Ile Val Gln Tyr Leu Asn Thr Thr Glu Gln Val Arg Ser Gly Asn
 145 150 155 160

Glu Ile Thr Cys Tyr Glu Asn Phe Thr Asp Asn Gln Leu Asp Val Val
 165 170 175
 Leu Pro Val Arg Leu Glu Leu Cys Leu Val Leu Phe Phe Ile Pro Met
 180 185 190
 5 Ala Val Thr Ile Phe Cys Tyr Trp Arg Phe Val Trp Ile Met Leu Ser
 195 200 205
 Gln Pro Leu Val Gly Ala Gln Arg Arg Arg Arg Ala Lys Gly Leu Ala
 210 215 220
 10 Val Val Thr Leu Leu Asn Phe Leu Val Cys Phe Gly Pro Tyr Asn Val
 225 230 235 240
 Ser His Leu Val Gly Tyr His Gln Arg Lys Ser Pro Trp Trp Arg Ser
 245 250 255
 Ile Ala Val Val Phe Ser Ser Leu Asn Ala Ser Leu Asp Pro Leu Leu
 260 265 270
 15 Phe Tyr Phe Ser Ser Ser Val Val Arg Arg Ala Phe Gly Arg Gly Leu
 275 280 285
 Gln Val Leu Arg Asn Gln Gly Ser Ser Leu Leu Gly Arg Arg Gly Lys
 290 295 300
 20 Asp Thr Ala Glu Gly Thr Asn Glu Asp Arg Gly Val Gly Gln Gly Glu
 305 310 315 320
 Gly Met Pro Ser Ser Asp Phe Thr Thr Glu
 325 330

(278) INFORMATION FOR SEQ ID NO:277:

- 25 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2724 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

- 30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:277:

ATGGACACCT CCCGGCTCGG TGTGCTCCTG TCCTTGCCTG TGCTGCTGCA GCTGGCGACC 60
 GGGGGCAGCT CTCCCAGGTC TGGTGTGTTG CTGAGGGGCT GCCCCACACA CTGTCATTGC 120
 GAGCCCGACG GCAGGATGTT GCTCAGGGTG GACTGCTCCG ACCTGGGGCT CTCGGAGCTG 180
 CCTTCCAACC TCAGCGTCTT CACCTCCTAC CTAGACCTCA GTATGAACAA CATCAGTCAG 240
 35 CTGCTCCCGA ATCCCCTGCC CAGTCTCCGC TTCCTGGAGG AGTTACGTCT TCGGGGAAAC 300

	GCTCTGACAT	ACATTCCCAA	GGGAGCATTC	ACTGGCCTTT	ACAGTCTTAA	AGTTCTTATG	360
	CTGCAGAATA	ATCAGCTAAG	ACACGTACCC	ACAGAAGCTC	TGCAGAATTT	GCGAAGCCTT	420
	CAATCCCTGC	GTCTGGATGC	TAACCACATC	AGCTATGTGC	CCCCAAGCTG	TTTCAGTGGC	480
	CTGCATTCCC	TGAGGCACCT	GTGGCTGGAT	GACAATGCGT	TAACAGAAAT	CCCCGTCCAG	540
5	GCTTTTAGAA	GTTTATCGGC	ATTGCAAGCC	ATGACCTTGG	CCCTGAACAA	AATACACCAC	600
	ATACCAGACT	ATGCCTTTGG	AAACCTCTCC	AGCTTGGTAG	TTCTACATCT	CCATAACAAT	660
	AGAATCCACT	CCCTGGGAAA	GAAATGCTTT	GATGGGCTCC	ACAGCCTAGA	GACTTTAGAT	720
	TTAAATTACA	ATAACCTTGA	TGAATTCCCC	ACTGCAATTA	GGACACTCTC	CAACCTTAAA	780
	GAAC TAGGAT	TTCATAGCAA	CAATATCAGG	TCGATACCTG	AGAAAGCATT	TGTAGGCAAC	840
10	CCTTCTCTTA	TTACAATACA	TTTCTATGAC	AATCCCATCC	AATTTGTTGG	GAGATCTGCT	900
	TTTCAACATT	TACCTGAACT	AAGAACACTG	ACTCTGAATG	GTGCCTCACA	AATAACTGAA	960
	TTTCCTGATT	TAAC TGGAAC	TGCAAACCTG	GAGAGTCTGA	CTTTAACTGG	AGCACAGATC	1020
	TCATCTCTTC	CTCAAACCGT	CTGCAATCAG	TTACCTAATC	TCCAAGTGCT	AGATCTGTCT	1080
	TACAACCTAT	TAGAAGATTT	ACCCAGTTTT	TCAGTCTGCC	AAAAGCTTCA	GAAAATTGAC	1140
15	CTAAGACATA	ATGAAATCTA	CGAAATTAAA	GTTGACACTT	TCCAGCAGTT	GCTTAGCCTC	1200
	CGATCGCTGA	ATTTGGCTTG	GAACAAAATT	GCTATTATTC	ACCCCAATGC	ATTTTCCACT	1260
	TTGCCATCCC	TAATAAAGCT	GGACCTATCG	TCCAACCTCC	TGTCGTCTTT	TCCTATAACT	1320
	GGGTTACATG	GTTTAACTCA	CTTAAATTA	ACAGGAAATC	ATGCCTTACA	GAGCTTGATA	1380
	TCATCTGAAA	ACTTTCCAGA	ACTCAAGGTT	ATAGAAATGC	CTTATGCTTA	CCAGTGCTGT	1440
20	GCATTTGGAG	TGTGTGAGAA	TGCCTATAAG	ATTTCTAATC	AATGGAATAA	AGGTGACAAC	1500
	AGCAGTATGG	ACGACCTTCA	TAAGAAAGAT	GCTGGAATGT	TTCAGGCTCA	AGATGAACGT	1560
	GACCTTGAAG	ATTTCTTGCT	TGACTTTGAG	GAAGACCTGA	AAGCCCTTCA	TTCAGTGCAG	1620
	TGTTACACCTT	CCCCAGGCC	CTTCAAACCC	TGTGAACACC	TGCTTGATGG	CTGGCTGATC	1680
	AGAATTGGAG	TGTGGACCAT	AGCAGTTCTG	GCACTTACTT	GTAATGCTTT	GGTGACTTCA	1740
25	ACAGTTTTCA	GATCCCCTCT	GTACATTTCC	CCCATTA AAC	TGTTAATTGG	GGTCATCGCA	1800
	GCAGTGAACA	TGCTCACGGG	AGTCTCCAGT	GCCGTGCTGG	CTGGTGTGGA	TGCGTTCACT	1860
	TTTGGCAGCT	TTGCACGACA	TGGTGCCTGG	TGGGAGAATG	GGGTTGGTTG	CCATGTCATT	1920
	GGTTTTTTGT	CCATTTTTGC	TTCAGAATCA	TCTGTTTTCC	TGCTTACTCT	GGCAGCCCTG	1980

GAGCGTGGGT TCTCTGTGAA ATATTCTGCA AAATTTGAAA CGAAAGCTCC ATTTTCTAGC 2040
 CTGAAAGTAA TCATTTTGCT CTGTGCCCTG CTGGCCTTGA CCATGGCCGC AGTTCCCCTG 2100
 CTGGGTGGCA GCAAGTATGG CGCCTCCCCT CTCTGCCTGC CTTTGCCTTT TGGGGAGCCC 2160
 AGCACCATGG GCTACATGGT CGCTCTCATC TTGCTCAATT CCCTTTGCTT CCTCATGATG 2220
 5 ACCATTGCCT ACACCAAGCT CTA CTGCAAT TTGGACAAGG GAGACCTGGA GAATATTTGG 2280
 GACTGCTCTA TGAAAAACA CATTGCCCTG TTGCTCTTCA CCAACTGCAT CCTAAACTGC 2340
 CCTGTGGCTT TCTTGTCTT CTCCTCTTTA ATAAACCTTA CATTTATCAG TCCTGAAGTA 2400
 ATTAAGTTTA TCCTTCTGGT GGTAGTCCCA CTTCTGCAT GTCTCAATCC CTTTCTCTAC 2460
 ATCTTGTTCA ATCCTCACTT TAAGGAGGAT CTGGTGAGCC TGAGAAAGCA AACCTACGTC 2520
 10 TGGACAAGAT CAAAACACCC AAGCTTGATG TCAATTAAT CTGATGATGT CGAAAAACAG 2580
 TCCTGTGACT CAACTCAAGC CTTGGTAACC TTTACCAGCT CCAGCATCAC TTATGACCTG 2640
 CCTCCAGTT CCGTGCCATC ACCAGCTTAT CCAGTGA CTG AGAGCTGCCA TCTTTCCTCT 2700
 GTGGCATTG TCCCATGTCT CTAA 2724

(279) INFORMATION FOR SEQ ID NO:278:

- 15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 907 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant
- 20 (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:278:
- | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Asp | Thr | Ser | Arg | Leu | Gly | Val | Leu | Leu | Ser | Leu | Pro | Val | Leu | Leu |
| 1 | | | | 5 | | | | 10 | | | | | | 15 | |
| Gln | Leu | Ala | Thr | Gly | Gly | Ser | Ser | Pro | Arg | Ser | Gly | Val | Leu | Leu | Arg |
| 20 | | | | | | | | 25 | | | | | 30 | | |
| Gly | Cys | Pro | Thr | His | Cys | His | Cys | Glu | Pro | Asp | Gly | Arg | Met | Leu | Leu |
| | 35 | | | | | | 40 | | | | | 45 | | | |
| Arg | Val | Asp | Cys | Ser | Asp | Leu | Gly | Leu | Ser | Glu | Leu | Pro | Ser | Asn | Leu |
| | 50 | | | | | 55 | | | | 60 | | | | | |
| Ser | Val | Phe | Thr | Ser | Tyr | Leu | Asp | Leu | Ser | Met | Asn | Asn | Ile | Ser | Gln |
| 65 | | | | | 70 | | | | | 75 | | | | 80 | |
| Leu | Leu | Pro | Asn | Pro | Leu | Pro | Ser | Leu | Arg | Phe | Leu | Glu | Glu | Leu | Arg |
| | | | | 85 | | | | | 90 | | | | | 95 | |

233

Leu Ala Gly Asn Ala Leu Thr Tyr Ile Pro Lys Gly Ala Phe Thr Gly
 100 105 110
 Leu Tyr Ser Leu Lys Val Leu Met Leu Gln Asn Asn Gln Leu Arg His
 115 120 125
 5 Val Pro Thr Glu Ala Leu Gln Asn Leu Arg Ser Leu Gln Ser Leu Arg
 130 135 140
 Leu Asp Ala Asn His Ile Ser Tyr Val Pro Pro Ser Cys Phe Ser Gly
 145 150 155 160
 10 Leu His Ser Leu Arg His Leu Trp Leu Asp Asp Asn Ala Leu Thr Glu
 165 170 175
 Ile Pro Val Gln Ala Phe Arg Ser Leu Ser Ala Leu Gln Ala Met Thr
 180 185 190
 Leu Ala Leu Asn Lys Ile His His Ile Pro Asp Tyr Ala Phe Gly Asn
 195 200 205
 15 Leu Ser Ser Leu Val Val Leu His Leu His Asn Asn Arg Ile His Ser
 210 215 220
 Leu Gly Lys Lys Cys Phe Asp Gly Leu His Ser Leu Glu Thr Leu Asp
 225 230 235 240
 20 Leu Asn Tyr Asn Asn Leu Asp Glu Phe Pro Thr Ala Ile Arg Thr Leu
 245 250 255
 Ser Asn Leu Lys Glu Leu Gly Phe His Ser Asn Asn Ile Arg Ser Ile
 260 265 270
 Pro Glu Lys Ala Phe Val Gly Asn Pro Ser Leu Ile Thr Ile His Phe
 275 280 285
 25 Tyr Asp Asn Pro Ile Gln Phe Val Gly Arg Ser Ala Phe Gln His Leu
 290 295 300
 Pro Glu Leu Arg Thr Leu Thr Leu Asn Gly Ala Ser Gln Ile Thr Glu
 305 310 315 320
 30 Phe Pro Asp Leu Thr Gly Thr Ala Asn Leu Glu Ser Leu Thr Leu Thr
 325 330 335
 Gly Ala Gln Ile Ser Ser Leu Pro Gln Thr Val Cys Asn Gln Leu Pro
 340 345 350
 Asn Leu Gln Val Leu Asp Leu Ser Tyr Asn Leu Leu Glu Asp Leu Pro
 355 360 365
 35 Ser Phe Ser Val Cys Gln Lys Leu Gln Lys Ile Asp Leu Arg His Asn
 370 375 380
 Glu Ile Tyr Glu Ile Lys Val Asp Thr Phe Gln Gln Leu Leu Ser Leu

	385	390	395	400
	Arg Ser Leu Asn Leu Ala Trp Asn Lys Ile Ala Ile Ile His Pro Asn	405	410	415
5	Ala Phe Ser Thr Leu Pro Ser Leu Ile Lys Leu Asp Leu Ser Ser Asn	420	425	430
	Leu Leu Ser Ser Phe Pro Ile Thr Gly Leu His Gly Leu Thr His Leu	435	440	445
	Lys Leu Thr Gly Asn His Ala Leu Gln Ser Leu Ile Ser Ser Glu Asn	450	455	460
10	Phe Pro Glu Leu Lys Val Ile Glu Met Pro Tyr Ala Tyr Gln Cys Cys	465	470	475
	Ala Phe Gly Val Cys Glu Asn Ala Tyr Lys Ile Ser Asn Gln Trp Asn	485	490	495
15	Lys Gly Asp Asn Ser Ser Met Asp Asp Leu His Lys Lys Asp Ala Gly	500	505	510
	Met Phe Gln Ala Gln Asp Glu Arg Asp Leu Glu Asp Phe Leu Leu Asp	515	520	525
	Phe Glu Glu Asp Leu Lys Ala Leu His Ser Val Gln Cys Ser Pro Ser	530	535	540
20	Pro Gly Pro Phe Lys Pro Cys Glu His Leu Leu Asp Gly Trp Leu Ile	545	550	555
	Arg Ile Gly Val Trp Thr Ile Ala Val Leu Ala Leu Thr Cys Asn Ala	565	570	575
25	Leu Val Thr Ser Thr Val Phe Arg Ser Pro Leu Tyr Ile Ser Pro Ile	580	585	590
	Lys Leu Leu Ile Gly Val Ile Ala Ala Val Asn Met Leu Thr Gly Val	595	600	605
	Ser Ser Ala Val Leu Ala Gly Val Asp Ala Phe Thr Phe Gly Ser Phe	610	615	620
30	Ala Arg His Gly Ala Trp Trp Glu Asn Gly Val Gly Cys His Val Ile	625	630	635
	Gly Phe Leu Ser Ile Phe Ala Ser Glu Ser Ser Val Phe Leu Leu Thr	645	650	655
35	Leu Ala Ala Leu Glu Arg Gly Phe Ser Val Lys Tyr Ser Ala Lys Phe	660	665	670
	Glu Thr Lys Ala Pro Phe Ser Ser Leu Lys Val Ile Ile Leu Leu Cys	675	680	685

235

Ala Leu Leu Ala Leu Thr Met Ala Ala Val Pro Leu Leu Gly Gly Ser
690 695 700

Lys Tyr Gly Ala Ser Pro Leu Cys Leu Pro Leu Pro Phe Gly Glu Pro
705 710 715 720

5 Ser Thr Met Gly Tyr Met Val Ala Leu Ile Leu Leu Asn Ser Leu Cys
725 730 735

Phe Leu Met Met Thr Ile Ala Tyr Thr Lys Leu Tyr Cys Asn Leu Asp
740 745 750

10 Lys Gly Asp Leu Glu Asn Ile Trp Asp Cys Ser Met Lys Lys His Ile
755 760 765

Ala Leu Leu Leu Phe Thr Asn Cys Ile Leu Asn Cys Pro Val Ala Phe
770 775 780

Leu Ser Phe Ser Ser Leu Ile Asn Leu Thr Phe Ile Ser Pro Glu Val
785 790 795 800

15 Ile Lys Phe Ile Leu Leu Val Val Val Pro Leu Pro Ala Cys Leu Asn
805 810 815

Pro Leu Leu Tyr Ile Leu Phe Asn Pro His Phe Lys Glu Asp Leu Val
820 825 830

20 Ser Leu Arg Lys Gln Thr Tyr Val Trp Thr Arg Ser Lys His Pro Ser
835 840 845

Leu Met Ser Ile Asn Ser Asp Asp Val Glu Lys Gln Ser Cys Asp Ser
850 855 860

Thr Gln Ala Leu Val Thr Phe Thr Ser Ser Ser Ile Thr Tyr Asp Leu
865 870 875 880

25 Pro Pro Ser Ser Val Pro Ser Pro Ala Tyr Pro Val Thr Glu Ser Cys
885 890 895

His Leu Ser Ser Val Ala Phe Val Pro Cys Leu
900 905

(280) INFORMATION FOR SEQ ID NO:279:

30 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

35 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:279:

CATGCCAACC GGCCCGCGAG GCTGCTGCTG GT

(281) INFORMATION FOR SEQ ID NO:280:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 32 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:280:

ACCAGCAGCA GCCTCGCGGG CCGGTTGGCA TG

32

INTERNATIONAL SEARCH REPORT

Int. Patent Application No.

PCT/US 99/23938

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/12 C07K14/72 G01N33/50 G01N33/566

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C07K G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	KJELSBERG M. A. ET AL.: "CONSTITUTIVE ACTIVATION OF THE ALPHA1B-ADRENERGIC RECEPTOR BY ALL AMINO ACID SUBSTITUTIONS AT A SINGLE SITE" JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 267, no. 3, 25 January 1992 (1992-01-25), pages 1430-1433, XP002911764 ISSN: 0021-9258 the whole document ---	1,2, 4-13, 15-33, 35-37,41
	-/--	

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

2 March 2000

Date of mailing of the international search report

09/03/2000

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Mandl, B

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SCHEER A. ET AL.: "CONSTITUTIVELY ACTIVE G PROTEIN-COUPLED RECEPTORS: POTENTIAL MECHANISMS OF RECEPTOR ACTIVATION" JOURNAL OF RECEPTOR AND SIGNAL TRANSDUCTION RESEARCH, vol. 17, no. 1/03, 1997, pages 57-73, XP000867531 ISSN: 1079-9893 the whole document	1,2, 4-13, 15-33, 35-37,41
X	WO 97 21731 A (NEW ENGLAND MEDICAL CENTER INC) 19 June 1997 (1997-06-19) the whole document, especially Fig. 2-3	1,2,4, 9-13, 20-32, 35-37,41
X	WO 98 38217 A (HERRICK DAVIS KATHARINE ;TEITLER MILT (US); EGAN CHRISTINA C (US)) 3 September 1998 (1998-09-03) the whole document, especially page 7, lines 24-27, and figure 4	1,2, 4-13, 15-33, 35-37,41
P,X	PAUWELS P. J. ET AL.: "REVIEW: AMINO ACID DOMAINS INVOLVED IN CONSTITUTIVE ACTIVATION OF G-PROTEIN-COUPLED RECEPTORS" MOLECULAR NEUROBIOLOGY, vol. 17, no. 1/03, 1998, pages 109-135, XP000866477 ISSN: 0893-7648 the whole document	1,2, 4-13, 15-33, 35-37,41

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 99/ 23938

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☒ Claims Nos.: 34, 38-40, 42, 43
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
See FURTHER INFORMATION sheet PCT/ISA/210
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US 99 /23938

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 34,38-40,42,43

Claims 34, 38-40, 42 and 43 refer to compounds with an agonistic effect on a GPCR without giving a true technical characterization. Moreover, no such specific compounds are defined in the application. In consequence, the scope of said claims is ambiguous and vague, and their subject-matter is not sufficiently disclosed and supported (Art. 5 and 6 PCT). No search can be carried out for such purely speculative claims whose wording is, in fact, a mere recitation of the results to be achieved.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 99/23938

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9721731 A	19-06-1997	US 5750353 A	12-05-1998
		AU 1334397 A	03-07-1997
		CA 2239293 A	19-06-1997
		EP 0869975 A	14-10-1998
WO 9838217 A	03-09-1998	AU 6343998 A	18-09-1998

THIS PAGE BLANK (USPTO)